

Research and Developments of Distributed Video Coding

A thesis submitted for the degree of Doctor of Philosophy

By
Zhuo Xue

School of Engineering and Design, Brunel University
January 2009

This work is dedicated to my parents, sister, and especially to
my wife who had endlessly supported and encouraged me
during this period of my three years study.

Acknowledgements

This thesis could not be finished without the help and support of many people who are gratefully acknowledged here.

Firstly, it is difficult to overstate my gratitude to my principle supervisor, Dr. Jonathan Loo. Without his guidance, I would not be able to finish this thesis finally. Throughout all my PhD study period, he provided encouragement, valuable advices, suggestions, criticisms and lots of novel academic ideas. His patience and kindness are greatly appreciated.

Secondly, I would like to thank Prof. John Cosmas who was my second supervisor, whose suggestions and encouragement helped me in all the time of research and writing of this thesis.

Thirdly, I would like to take this opportunity to express my gratitude to Prof. Jian Min Jiang and Dr Maysam Abbot who are my external examiner and internal examiner. Their valuable suggestions and comments help improve the quality of this thesis significantly.

What's more, I would like to gratefully acknowledge the financial support from the ORSAS and the School of Engineering and Design, Brunel University, UK. Without this support, I wouldn't be able to fully concentrate on my study. I also want to take this opportunity to thank Dr. Myo Tun who has given me valuable instructions to use the DIRAC software.

Finally, I would like to give my special thanks to my wife Dr. Lifang Feng whose endless love supported me to complete this work.

Abstract

The recent developed Distributed Video Coding (DVC) is typically suitable for the applications such as wireless/wired video sensor network, mobile camera etc. where the traditional video coding standard is not feasible due to the constrained computation at the encoder. With DVC, the computational burden is moved from encoder to decoder. The compression efficiency is achieved via joint decoding at the decoder.

The practical application of DVC is referred to Wyner-Ziv video coding (WZ) where the side information is available at the decoder to perform joint decoding. This joint decoding inevitably causes a very complex decoder. In current WZ video coding issues, many of them emphasise how to improve the system coding performance but neglect the huge complexity caused at the decoder. The complexity of the decoder has direct influence to the system output. The beginning period of this research targets to optimise the decoder in pixel domain WZ video coding (PDWZ), while still achieves similar compression performance. More specifically, four issues are raised to optimise the input block size, the side information generation, the side information refinement process and the feedback channel respectively.

The transform domain WZ video coding (TDWZ) has distinct superior performance to the normal PDWZ due to the exploitation in spatial direction during the encoding. However, since there is no motion estimation at the encoder in WZ video coding, the temporal correlation is not exploited at all at the encoder in all current WZ video coding issues. In the middle period of this research, the 3D DCT is adopted in the TDWZ to remove redundancy in both spatial and temporal direction thus to provide even higher coding performance. In the next step of this research, the performance of transform domain Distributed Multiview Video Coding (DMVC) is also investigated. Particularly, three types transform domain DMVC frameworks which are transform domain DMVC using TDWZ based 2D DCT, transform domain DMVC using TDWZ based on 3D DCT and transform domain residual DMVC using TDWZ based on 3D DCT are investigated respectively.

One of the important applications of WZ coding principle is error-resilience. There have been several attempts to apply WZ error-resilient coding for current video coding standard e.g. H.264/AVC or MPEG 2. The final stage of this research is the design of WZ error-resilient scheme for wavelet based video codec. To balance the trade-off between error resilience ability and bandwidth consumption, the proposed scheme emphasises the protection of the Region of Interest (ROI) area. The efficiency of bandwidth utilisation is achieved by mutual efforts of WZ coding and sacrificing the quality of unimportant area.

In summary, this research work contributed to achieves several advances in WZ video coding. First of all, it is targeting to build an efficient PDWZ with optimised decoder. Secondly, it aims to build an advanced TDWZ based on 3D DCT, which then is applied into

multiview video coding to realise advanced transform domain DMVC. Finally, it aims to design an efficient error-resilient scheme for wavelet video codec, with which the trade-off between bandwidth consumption and error-resilience can be better balanced.

Table of Contents

1. Introduction.....	18
1.1 Background and Motivation	19
1.2 Aim and Objectives	20
1.3 Main Contributions.....	20
1.4 Organisation of the Thesis	21
2. Review of Wyner-Ziv Video Coding.....	23
2.1 Theoretic Background.....	23
2.1.1 Slepian-Wolf Theorem.....	24
2.1.2 Wyner-Ziv Theorem.....	25
2.1.3 Practical Wyner-Ziv Code Design	26
2.2 Wyner-Ziv Video Coding-Stanford University Solution	28
2.2.1 Overall Architecture.....	28
2.2.2 Quantisation	31
2.2.3 Transformation.....	31
2.2.4 Slepian-Wolf Encoder.....	32
2.2.5 Feedback Channel.....	33
2.2.6 Side Information	34
2.2.7 Slepian-Wolf Decoder.....	35
2.2.8 Frame Reconstruction	38
2.2.9 Experimental Results	39
2.3 Wyner-Ziv video coding- Instituto Superior Técnico Solution.....	42
2.3.1 IST PDWZ	42
2.3.2 IST TDWZ	45
2.3.3 Experimental Results	45
2.4 Wyner-Ziv Video Coding-PRISM Solution.....	46
2.4.1 Encoding	47
2.4.2 Decoding.....	48
2.4.3 Experimental Results	49
2.5 Other Advances in Wyner-Ziv Video Coding	49

2.5.1	Advances in Quantisation	49
2.5.2	Advances in Transformation.....	50
2.5.3	Advances in Slepian-Wolf Codec	50
2.5.4	Advances in Modelling Correlation Channel.....	51
2.5.5	Advances in Side Information Generation.....	52
2.5.6	Advances in Other Aspects	54
2.6	Further Discussions.....	54
2.7	Chapter Summary	55
3.	Pixel Domain Wyner-Ziv Video Coding with Optimised Decoder	56
3.1	Slice Structure.....	56
3.1.1	Background and Motivation	56
3.1.2	Architecture and Operations	58
3.1.3	Results and Discussions.....	60
3.2	Backward Multiple Reference Frame Interpolation.....	63
3.2.1	Review of Frame Interpolation	63
3.2.2	BMRF	64
3.2.3	Results and Discussions.....	67
3.3	Embedded Side Information Refinement.....	71
3.3.1	Review of Side Information Refinement	71
3.3.2	Architecture and Operations	72
3.3.3	Results and Discussions.....	74
3.4	Adaptive feedback optimisation	77
3.4.1	Review of Rate Control	78
3.4.2	Algorithm and Operations.....	79
3.4.3	Results and Discussions.....	82
3.5	Chapter Summary	84
4.	Transform Domain WZ Video Coding with 3D DCT and Applications	87
4.1	Background and Motivation	87
4.2	3D DCT	88
4.3	Transform Domain WZ Video Coding with 3D DCT.....	91
4.3.1	Architecture and Operations	91

4.3.2	WZ Frame Grouping (WFG)	93
4.3.3	Side Information Grouping (SIG)	94
4.3.4	Quantisation Volume Design	97
4.3.5	Results and Discussions.....	100
4.4	Transform Domain Distributed Multiview Video Coding	103
4.4.1	Review of DMVC	103
4.4.2	Configuration in DMVC	106
4.4.3	Pixel Domain DMVC	109
4.4.4	Transform Domain DMVC with 2D DCT	111
4.4.5	Transform Domain DMVC with 3D DCT	113
4.4.6	Transform Domain Residual DMVC with 3D DCT	114
4.4.7	Results and Discussions.....	116
4.5	Chapter Summary	118
5.	Efficient Error Resilient Scheme for Wavelet Video Codec using Wyner-Ziv Coding	120
5.1	Review of Error-Resilient Techniques for Wavelet Video Codec.....	120
5.2	Automatic ROI Detection	123
5.2.1	Frame Difference Mask Definition	124
5.2.2	ROI Generation.....	124
5.2.3	D_{th} Decision	124
5.3	ROI Mask Generation in Wavelet Domain	126
5.4	Receiver Driven Layered Wyner-Ziv Coding.....	128
5.5	WZER for Packet Erasure Channel	129
5.5.1	Architecture	130
5.5.2	Packet Level LDPC	131
5.5.3	Results and Discussions.....	133
5.6	WZER for AWGN Channel	141
5.6.1	Architecture	141
5.6.2	Simulation Results	143
5.7	Chapter Summary	150
6.	Conclusions and Future Work.....	152
6.1	Overall Achievements.....	152

6.1.1	Pixel domain WZ Video Coding with Optimised Decoder.....	152
6.1.2	3D-TDWZ and its Application in Multiview Video Coding.....	153
6.1.3	Error Resilient Schemes for Wavelet Video Codec using WZ Coding	154
6.2	Future Works.....	155
6.2.1	Slice Structure.....	155
6.2.2	Rate Control in WZ Video Coding	155
6.2.3	Error Resilience with WZ Coding	156

List of Figures

Fig.2. 1	Architecture of independent encoding and independent decoding for two sources	24
Fig.2. 2	Architecture of joint encoding and joint decoding for two sources.....	24
Fig.2. 3	Architecture of independent encoding and joint decoding for two sources.....	24
Fig.2. 4	Achievable rate region of Slepian-Wolf theorem [3].....	25
Fig.2. 5	Wyner-Ziv theorem: lossy compression with side information.....	26
Fig.2. 6	STD pixel domain Wyner-Ziv video coding [3, 25].....	28
Fig.2. 7	STD transform domain Wyner-Ziv Video coding [26].....	29
Fig.2. 8	Quantisation matrices used for transform domain WZ coding [26]	31
Fig.2. 9	Coefficient bands grouping in TDWZ	32
Fig.2. 10	The encoder architecture of turbo encoder	33
Fig.2. 11	Example puncturing pattern for turbo encoding	33
Fig.2. 12	SISO decoding in Turbo decoder [28]	35
Fig.2. 13	The Laplacian prediction process of soft input of the decoder.....	36
Fig.2. 14	Frame reconstruction function (4 quantisation levels) [25].....	38
Fig.2. 15	RD performance of STD WZ codec with Salesman sequence [3]	39
Fig.2. 16	RD performance of STD WZ codec with Hall sequences [3]	40
Fig.2. 17	RD performance of STD TDWZ with different length of GOP for Salesman [3]... <td>41</td>	41
Fig.2. 18	RD performance of STD TDWZ with different length of GOP for Hall [3]	41
Fig.2. 19	The architecture of IST PDWZ [33]	43
Fig.2. 20	Side information generation framework for IST PDWZ[33]	43
Fig.2. 21	Forward ME for side information interpolation in IST PDWZ[33]	44
Fig.2. 22	Bidirectional ME for side information interpolation in IST PDWZ[33]	44
Fig.2. 23	Perceptual comparison for side information with or without spatial smoothing [33]	44
Fig.2. 24	PSNR comparison for IST and STD WZ codec with Foreman sequence@30HZ ..	45
Fig.2. 25	The architecture of PRISM [37]	46
Fig.2. 26	Bit stream syntax in PRISM[37]	48
Fig.2. 27	RD performance of STD TDWZ with different side information interpolations for Foreman [31].....	52
Fig.3. 1	The architecture of proposed PDWZ with Slice structure	58
Fig.3. 2	PSNR performance of PDWZS for Foreman sequence.....	61
Fig.3. 3	PSNR performance of PDWZS for Mother and daughter sequence	61
Fig.3. 4	Perceptual performance of PDWZS and normal PDWZ	62
Fig.3. 5	Average time (seconds) consumed comparison between PDWZS and normal PDWZ	62
Fig.3. 6	Weighted Frame interpolation in BMRF	65
Fig.3. 7	PSNR comparison for TLIN, BID and BMRF interpolation for Foreman sequence	67
Fig.3. 8	PSNR comparision for TLIN,BID and BMRF interpolation for Carphone sequence	68
Fig.3. 9	PSNR comparison for TLIN,BID and BMRF interpolation for Coastguard sequence	68
Fig.3. 10	PSNR comparison for TLIN,BID and BMRF interpolation for Salesman sequence	69
Fig.3. 11	Perceptual quality comparison for TLIN, BID and BMRF interpolation.....	70

Fig.3. 12 The architecture of PDWZS with embedded refinement.....	73
Fig.3. 13 Embedded refinement inside the turbo decoder	74
Fig.3. 14 PSNR performance of PDWZS +BMRF +WMCR refinement for Foreman.....	75
Fig.3. 15 PSNR performance of PDWZS+BMRF+WMCR refinement for Carphone	76
Fig.3. 16 A typical residual distribution in WZ video coding.....	80
Fig.3. 17 Probability distribution of final decoding code rate with different range of	82
Fig.3. 18 Perceptual comparison with different rate control algorithms.....	84
Fig.4. 1 The architecture of transform domain DVC with 3D DCT	91
Fig.4. 2 WFG for $n=2$	93
Fig.4. 3 WFG for $n>2$ and $l=n-1$	93
Fig.4. 4 WFG for $n>2$ and $l=n$	94
Fig.4. 5 Side information Interpolation for $n=2$, $l=3$	95
Fig.4. 6 Side information interpolation for $n=3, l=2$	95
Fig.4. 7 Side information interpolation with $n=4$, $l=3$	96
Fig.4. 8 Side information interpolation for $n>2$, $l \geq n$	96
Fig.4. 9 Distribution of dominant coefficients in DCT coefficient cube	97
Fig.4. 10 Shape of the distribution of the dominant coefficients[109]	98
Fig.4. 11 R-D performance of 3D-TDWZ with GOP length $n=2$ of Mother and daughter sequence	101
Fig.4. 12 R-D performance comparison for 3D-TDWZ for Hall sequence	102
Fig.4. 13 R-D performance comparison for 3D-TDWZ for Salesman sequence.....	102
Fig.4. 14 Multiview video coding case 1(conventional coding).....	104
Fig.4. 15 Multiview video coding case 2(Interview coding)	104
Fig.4. 16 A typical interview video coding structure [118].....	104
Fig.4. 17 Multiview video coding case 3(Distributed multiview video coding)	105
Fig.4. 18 Frame composition of Key camera and WZ camera in [100].....	106
Fig.4. 19 Frame composition of Key camera and WZ camera in [36].....	106
Fig.4. 20 Frame composition of Key camera and WZ camera in [103].....	107
Fig.4. 21 The architecture of joint side information generation	108
Fig.4. 22 Side information generation in JSG.....	108
Fig.4. 23 Side information comparison for different interpolation in multiveiw.....	109
Fig.4. 24 The architecture of pixel domain DMVC.....	110
Fig.4. 25 The architecture of transform domain DMVC with 2D DCT	112
Fig.4. 26 Quantisation matrix for transform domain DMVC with 2D DCT [31].....	113
Fig.4. 27 The architecture of transform domain DMVC with 3D DCT	113
Fig.4. 28 The architecture of Transform domain residual DMVC with 3D DCT	115
Fig.4. 29 R-D performance of proposed transform domain DMVC for Vassar sequence (Camera 0).....	117
Fig.4. 30 R-D performance of proposed transform domain DMVC for Ballroom sequence (Camera 1).....	117
Fig.5. 1 ROI area detection with different F_{th}	126
Fig.5. 2 Wavelet transform with lifting scheme.....	126
Fig.5. 3 ROI mask example	128
Fig.5. 4 Layer design of the proposed WZER scheme	129
Fig.5. 5 The proposed WZER for wavelet video coding over packet erasure channel.....	130
Fig.5. 6 The coding process of LDPC code [160]	132
Fig.5. 7 LDPC FEC scheme.....	134

Fig.5. 8	Average PSNR performance of WZER under packet erasure channel	136
Fig.5. 9	PSNR performance vs frame number of WZER	137
Fig.5. 10	Perceptual comparison of I, B, P frames with WZER of Akiyo sequence	138
Fig.5. 11	PSNR and perceptual performance for WZER with different F_{th}	140
Fig.5. 12	Perceptual comparison for WZER with different number of sub bands received	141
Fig.5. 13	The proposed WZER for wavelet video coding over AWGN channel	142
Fig.5. 14	PSNR performance of WZER for Akiyo sequence	145
Fig.5. 15	PSNR performance of WZER for Silent sequence	145
Fig.5. 16	Perceptual comparison of I, B, P frames of WZER for Akiyo sequence	146
Fig.5. 17	Perceptual comparison of I, B, P frames of WZER for Silent sequence	147
Fig.5. 18	Average PSNR performance of WZER with different F_{th} of Akiyo sequence.....	148
Fig.5. 19	Average PSNR performance of WZ stream with different Fth values of Akiyo sequence	148
Fig.5. 20	Perceptual quality comparison of WZER with different F_{th} of Akiyo sequence ...	149
Fig.5. 21	Perceptual comparison for WZER with different number of sub bands of Akiyo sequence	150

List of Tables

Table 3. 1 PSNR performance of different MCI algorithms	69
Table 3. 2 Computational complexity of different MCI algorithms	70
Table 3. 3 Monitored BER of embedded refinement for 4th slice of 1st WZ frame of ‘Carpone’ sequence with 16 quantisation level	77
Table 3. 4 PSNR between original 4th slice of 1st WZ frame and it’s side information and refined slice after each iteration(dB)	77
Table 3. 5 Proposed AFO scheme for puncturing period of 8.....	82
Table 3. 6 R-D performance comparison for AFO and ERC scheme	83
Table 4. 1 An example quantisation volume($C=15, \beta_i=0.001, \beta_o=0.03$)	99
Table 5. 1 Error-resilient schemes for Akiyo over packet erasure channel.....	134
Table 5. 2 Error-resilient schemes for Silent over packet erasure channel	134
Table 5. 3 Error-resilient schemes for Akiyo over AWGN channel.....	144
Table 5. 4 Error-resilient schemes for Silent over AWGN channel	144

List of Acronyms

1D	One Dimension
2D	Two Dimension
2D-TDWZ	Transform Domain Wyner-Ziv video coding based on 2D DCT
2TD-DMVC	Transform Domain Distributed Multiview Video Coding based on 3D DCT
3D	Three Dimension
3D-TDWZ	Transform Domain Wyner-Ziv video coding based on 3D DCT
3GPP	The 3rd Generation Partner Project
3TD-DMVC	Transform Domain Distributed Multiview Video Coding based on 3D DCT
3TDS-DMVC	Transform Domain Residual Distributed Multiview Video Coding based on 3D DCT
AFO	Adaptive feedback optimisation
ARQ	Automatic Repeat-reQuest
AVC	Advance Video Coding
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BID	Bidirectional motion compensated interpolation
BME	Block Motion Estimation
BMRF	Backward Multiple Reference Frame
CDF	Cohen-Daubechies-Feauveau
CIF	Common Intermediate Format
CRC	Cyclic Redundancy Check
DCT	Discrete Cosine Transform
DMVC	Distributed Multiview Video Coding
DSC	Distributed Source Coding
DISCOVER	Distributed Coding for Video Services
DVC	Distributed Video Coding
DWT	Discrete Wavelet Transform
ERC	Exhaustive Rate Control
EZW	Embedded Zerotrees Wavelet
FEC	Forward Error Correction
FMO	Flexible Macroblock Ordering
GOP	Group of Picture
GOW	Group of WZ frames
HDTV	Hi Definition Television
IDCT	Inverse Discrete Cosine Transform
IDWT	Inverse Discrete Wavelet Transform

IETF	Internet Engineering Task Force
i.i.d.	Independent Identically Distributed
IST	Instituto Superior Técnico
JDU	Joint Decoding Unit
JPEG	Joint Photographic Experts Group
JSG	Joint Side information Generation
LDPC	Low-Density Parity-Check
LVQ	Lattice Vector Quantisation
MAP	Maximum A Posteriori
MB	Micro Block
MC	Motion Compensation
MCE	Motion Compensated Extrapolation
MCI	Motion Compensated Interpolation
MCR	Motion Compensated Restoration
ME	Motion Estimation
MERL	Mitsubishi Electric Research Laboratories
ML	Maximum Likelihood
MPEG	Moving Pictures Expert Group
MSE	Mean Square Error
MSD	Mean Square Difference
MV	Motion Vector
OBME	Overlapped Block Motion Estimation
PDWZ	Pixel Domain Wyner-Ziv
PDWZS	Pixel Domain Wyner-Ziv video coding with Slice
PRISM	Power-efficient, Robust, High-compression, Syndrome-based Multimedia coding
PSNR	Peak Signal to Noise Ratio
QCIF	Quarter Common Intermediate Format
RAM	Random Access Memory
RCPC	Rate-Compatible Punctured Convolutional
RCPT	Rate Compatible Punctured Turbo codec
RD	Rate Distortion
RDO	Rate-Distortion Optimisation
ROI	Region Of Interest
RS	Reed-Solomon
RSC	Recursive Systematic Convolutional
SAD	Sum of Absolute Difference
SIF	Standard Image Format

SIG	Side Information Grouping
SISO	Soft-Input Soft-Output
SLEP	Systematic Lossy Error Protection
SNR	Signal to Noise Ratio
SPIHT	Set Partitioning in Hierarchical Trees
SRC	Selective Rate Control
STD	Stanford Research Group
SWSPIHT	Slepian-Wolf Set Partitioning in Hierarchical Trees
TDWZ	Transform Domain Wyner-Ziv
TLIN	Temporal Linear motion compensated interpolation
TMN8	Test Model Near-term Version 8
VLC	Variable Length Coding
WAMCI	Weighted-adaptive Motion-compensated Interpolation
WFG	WZ Frame Grouping
WMCR	Weighted Motion Compensation Refinement
WZ	Wyner-Ziv
WZER	Wyner-Ziv Error-Resilient
ZTE	Zerotree Entropy

List of Publication

Journals

- [1] Z. Xue, K. K. Loo, and J. Cosmas, "New Multiple Frame Motion Compensated Interpolation Algorithm for Frame Rate up-Conversion," *Journal of Circuits, Systems and Computers*, vol. 11, pp.835-844, 2008.
- [2] Z. Xue, K. K. Loo, J. Cosmas, and P. Y. Yip, "Distributed video coding in wireless multimedia sensor network for multimedia broadcasting," *WSEAS Transactions on Communications*, vol. 7, pp. 418-427, 2008.
- [3] Z. Xue, K. K. Loo, J. Cosmas, and P. Y. A. Yip, "An efficient error resilience scheme based on Wyner-Ziv coding for region-of-interest protection of wavelet based video transmission," *WSEAS Transactions on Circuits and Systems*, vol. 7, pp. 277-286, 2008..
- [4] Z. Xue, K. K. Loo, and J. Cosmas, M.Tun" Error-resilient Scheme for Wavelet Video Codec over packet erasure channel using automatic ROI detection and Wyner-Ziv coding," *IEEE Transaction on Broadcasting*, Submitted.
- [5] Z. Xue, K. K. Loo, and J. Cosmas, P. Y. A. Yip " Error-resilient Scheme for Wavelet Video Codec using packet level RS code and Wyner-Ziv coding," *International Journal of Wavelets, Multiresolution and Information Processing*, Submitted.
- [6] Z. Xue, K. K. Loo, and J. Cosmas, "Transform domain Wyner-Ziv video coding with 3D DCT," *IEEE Signal Processing letter*, Submitted.
- [7] Z. Xue, K. K. Loo, and J. Cosmas, "Transform domain Distributed Multiview Video coding," *Springer Journal of Signal Processing Systems*, Submitted.
- [8] Z. Xue, K. K. Loo, and J. Cosmas, "Advanced transform domain Wyner-Ziv coding with 3D DCT in Distributed Multiview Video coding," *Springer Journal of Signal, Image and Video Processing*, Submitted
- [9] Z. Xue, K. K. Loo, and J. Cosmas, "Error-resilient scheme using automatic ROI detection and receiver-driven layered Wyner-Ziv protection for wavelet codec transmission in AWGN channel," *Elsevier Journal of Signal Processing: Image Communication*, Submitted

Conferences

- [1] Z. Xue and K. K. Loo, "New side information interpolation with multiple reference frame MCI in distributed video coding," *IEEE International Conference on Consumer Electronics(ICCE 07)*,2007, Las Vegas, USA.
- [2] Z. Xue, K. K. Loo, and J. Cosmas, "Embedded side information refinement for pixel domain Wyner-Ziv video coding towards UMTS 3G application," *IEEE International Conference on Multimedia and Expo (ICME 2007)*, 2007, pp. 156-159, Beijing, P.R.China
- [3] Z. Xue, K. K. Loo and J. Cosmas, "Wireless Multimedia Sensor Network with an Efficient Distributed Video Coding for Multimedia Broadcasting," *8th WSEAS Int. Conf. on Multimedia Systems and Signal Processing (MUSP '08)*, pp 273-279, April 6-8, 2008, Hangzhou, China.
- [4] Z. Xue, K. K. Loo and J. Cosmas, "Error Resilient Wavelet Video Transmission with Priority area Protection using Wyner-Ziv Coding," *8th WSEAS Int. Conf. on Multimedia Systems and Signal Processing (MUSP '08)*, pp 273-279, April 6-8, 2008, Hangzhou, China.
- [5] Z. Xue, K. K. Loo, and J. Cosmas, "Bandwidth efficient error resilience scheme for wavelet based video transmission," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2008*, Las Vegas, Nevada, USA.

Chapter 1

1. Introduction

In current international digital video compression standards such as MPEG-4 or H.264/AVC, the inter-frame predictive coding is used to exploit the redundancy among successive frames to achieve high compression. However, the utilisation of motion estimation (ME) in the predictive coding inevitably causes an asymmetry video codec which has a more complex encoder and relatively simpler decoder. The video encoder is typically 5 to 10 times more complex than the decoder. This asymmetry structure is suitable for storage purpose or the applications of broadcasting or streaming video-on-demand systems where video is compressed once and decoded for many times. However, in certain systems, an inverse architecture is required where the encoder is constrained in the computation while decoder can afford more complex tasks. The typical examples of such applications include wireless/wired sensor network, mobile camera video processing, wireless PC camera and video surveillance system etc where the computation and memory at the encoder are scarce. Due to the energy constraint of the encoder, the predictive coding is not suitable for such applications.

The applications with low complexity encoder still need to achieve compression efficiency without predictive coding. The research achievements in distributed source coding theorem [1, 2] suggested that efficient compression can be achieved by exploiting source statistics partially or wholly at the decoder only. These theorems are referred as Slepian-Wolf theorem for distributed lossless coding and Wyner-Ziv (WZ) theorem for distributed lossy coding with side information at the decoder. Based upon these two theorems, distributed video coding (DVC) devotes to offering the solutions for above suggested architectures has been developed in last few years. Particularly, Wyner-Ziv video coding, a practical case of DVC based on applying WZ theorem in real video coding has been extensively studied.

The most attractive advantage of WZ video coding algorithm is that it moves the computation burden from the encoder to the decoder but the compression efficiency still can be achieved by performing joint decoding at the decoder. Although numerous research achievements around WZ video coding have been made in last few years, the compression performances of WZ video coding still cannot match predictive video coding so far. DVC is still far from mature to be commercialised. There is still a wide space in the DVC field that needs to be explored in the future.

1.1 Background and Motivation

WZ video coding has advantages in several applications due to moving the burden of the computation from the encoder to the decoder. However, since WZ video coding has no restriction on the decoder complexity, most current WZ video coding issues are regarding to how to improve the system performance whilst neglect the negative influence caused by complicate decoder. The complexity of the decoder cannot be infinitely large and it is directly related to the decoding efficiency and delay of the system output. In current WZ video coding algorithms, there are some inefficient operations which cause a very complicated decoder. Some assumptions in the WZ video coding are not realistic such as the input block with arbitrary size, repetitive side information refinement, or feedback channel which allows innumerable retransmission and re-decoding etc. In fact these operations can be further optimised thus to improve the decoding efficiency while the system performance still can be maintained similarly.

Moreover, traditional video coding standards such as H.26X and MPEG-X achieve high compression via removing the redundancy of spatial and temporal direction by spatial transformation and ME. One of the reasons that WZ video coding cannot match the current predictive video coding is that the temporal correlation of WZ video coding is not fully exploited. In traditional video coding standards, ME and Motion Compensation (MC) are used to achieve this goal. But ME and MC are not feasible in WZ video coding because of the constraint in the complexity at the encoder. However, apart from predictive coding strategy, there are other options available such as 3D DCT that can be used to realise the temporal compression and keep the encoder low complexity meanwhile. In current WZ video coding, compression performance is improved in transform domain WZ video coding architecture (TDWZ). It is proved TDWZ is superior to the normal pixel domain WZ video coding (PDWZ) due to the transformation used to exploit the spatial compression. The compression performance of TDWZ can still be further enhanced by using 3D DCT to remove the temporal redundancy.

One of the most popular applications of WZ video coding is multiview video coding. The advantage of applying WZ coding principle into multiview video coding is that this new architecture called Distributed multiview video coding (DMVC) can offer the solution to the multiview video coding applications with limited resources at the encoder. In DMVC, the video content of each camera is separately encoded but jointly decoded at the decoder. Due to the correlation exploitation done at the encoder, the compression efficiency is achieved. However, in most DMVC architectures, PDWZ is widely applied and TDWZ is seldom explored. Since the TDWZ can provide better compression performance than PDWZ coding

while the complexity of the encoder still keeps similar as that in intra-frame coding, to apply TDWZ video coding in DMVC to compose a transform domain DMVC is worth studying.

Another important application of WZ coding is robust video transmission. It is presented in many references that using WZ coding principle for error-resilience can provide competitive error resilient performance comparing to the traditional FEC coding. Most available references consider the error-resilience using WZ coding applied in traditional video coding standard e.g. H.264 or mpeg 2. Since video codec based on wavelet transformation is increasingly popular and it is necessary to propose error-resilient scheme using WZ coding technique for such wavelet video codec. How to design such error-resilient scheme for this kind of video codec based on wavelet remains unexplored. How to design an efficient WZ error-resilient coding method to compete the traditional FEC coding and how does it perform need to be addressed.

1.2 Aim and Objectives

The main goal of this research is to develop an advanced and efficient WZ coding architecture and apply it into related applications. A series of researches had been conducted to achieve this goal.

Firstly, in order to optimise the decoder efficiency, several issues are raised aiming to adjust WZ video coding structure in the aspect of changing input block size, changing side information interpolation method, adjusting side information refinement techniques and optimising the feedback process.

Secondly, 3D DCT transformation techniques are applied in WZ video coding to exploit the temporal redundancy thus to enhance overall system performance. Afterwards, the designed TDWZ is applied in the multiview video coding to constitute advanced transform domain DMVC.

Finally, a WZ error-resilient scheme is designed for wavelet based video codec. The error-resilience and bandwidth usage are considered equally in the proposed algorithm to achieve better balance between them. The performance is compared with traditional FEC coding algorithm and verified under different channel type.

1.3 Main Contributions

In the beginning phase of the research, all works are devoted to optimising the decoder in WZ video coding. Firstly, a slice structure is used as the way to decompose the WZ frame for transmission. The slice structure can reduce the input block size and make it compatible to the practical channel code. Not only it effectively reduces the decoding complexity, but also it optimises the rate control process inside the WZ video coding thus the system enhancement is obtained. To further reduce unnecessary complicate operations at the decoder, a fast

Backward Multiple Reference Frame (BMRF) interpolation technique is proposed to replace the widely used bidirectional motion compensated interpolation method which is considered as an exhaustive and time consuming process. In addition, an embedded side information refinement technique is proposed in order to avoid repetitive turbo decoding which is usually happened in the current side information refinement techniques. By cooperating with the BMRF, the refinement technique also compensates the quality lost of the side information during BMRF process and improves the system performance eventually. The final work targeting to reduce the decoding complexity is an adaptive feedback optimisation scheme which optimises the feedback process in the WZ video coding architecture.

In the middle phase of the research, an advanced TDWZ based on 3D DCT (3D-TDWZ) is proposed. The 3D DCT used is to exploit the temporal correlation while still keeps the encoder at low complexity. The proposed 3D-TDWZ especially is suitable for working with larger GOP(Group of pictures) and the performance is verified via simulation. Subsequently, the proposed 3D-TDWZ and TDWZ based 2D DCT (2D-TDWZ) are applied in multiview video coding to constitute advanced transform domain DMVC. In total three types of transform domain DMVC architectures are proposed and compared to the normal pixel domain DMVC suggested by current available references.

In the final phase of the research, an error-resilient scheme called WZER is specially designed for wavelet based video codec by using WZ coding. More importantly, WZER is designed to achieve better trade-off between bandwidth usage and error-resilience by emphasizing the protection of Region of Interest (ROI) area other than the whole frame. The performances are verified by comparing WZER with traditional FEC scheme over packet erasure channel and AWGN channel.

1.4 Organisation of the Thesis

The thesis is composed of six chapters where the first chapter is introduction followed by the detail review of WZ video coding in Chapter 2. In Chapter 3, the main contributions targeting to design efficient and advanced WZ video coding architecture is discussed. Subsequently, the TDWZ architecture using 3D DCT and transform domain DMVC is proposed in Chapter 4. Chapter 5 contains the detail design of the proposed WZER scheme and corresponding simulations. Final conclusions of this thesis and future works recommended are discussed in Chapter 6

Chapter 1: Starting with a brief introduction of current mature international video coding standards, some important applications are highlighted with which most standard video coding schemes are not capable of dealing. Then DVC is presented and the distinct advantages are listed comparing to the current international video coding standards. In addition, background and motivation, aims and objectives, main contributions of the research

were presented in details. Finally, the organisation of the thesis is stated with the overall description of each chapter.

Chapter 2: The review of current WZ video coding is presented. Firstly, the background theory support of DVC is introduced. Then the WZ video coding architectures from several leading research groups are presented subsequently, where each key component and details of operation together with experimental results are included. Moreover, some advances in other aspects presented in research community are reviewed in different categories.

Chapter 3: This chapter mainly devotes to optimising the decoder efficiency and offering a low-delay system. These issues include advanced Pixel Domain Wyner-Ziv video coding with slice structure (PDWZS), the algorithm design of BMRF, the embedded refinement technique of side information and the adaptive feedback optimisation (AFO) scheme.

Chapter 4: The design of transform domain WZ video coding based on 3D DCT and transform domain DMVC are presented. The chapter begins with a brief review of transform domain WZ video coding and the motivations. After that, the transform domain DVC with 3D DCT is presented and the design of transform domain DMVC is described in later sections. Results and discussion are drawn after each scheme respectively.

Chapter 5: Some WZER schemes for wavelet based video codec are described. The chapter begins with a detail review of current error resilient scheme for wavelet video transmission and brings the motivation of the research. Several key components used to build up the WZER scheme are illustrated including automatic ROI detection, ROI mask generation, and packet level FEC protection, and layered WZ coding design etc. Then the WZER architecture is proposed and followed by simulation results over different error prone channels finally.

Chapter 6: This chapter concludes the whole thesis where overall achievements of this research and further recommendations of each phase of the research are presented.

Chapter 2

2. Review of Wyner-Ziv Video Coding

This chapter presents a general review of WZ video coding from theory background to the recent advances in WZ video coding filed. Firstly, the basic theory of Distributed Source Coding (DSC) that WZ video coding relies on is introduced, in which Slepian-Wolf and Wyner-Ziv theorems are explained in details. There are several leading research groups in DVC field. Section 2.2 presents the review of a typical WZ video coding solution proposed by Stanford (STD) university DVC research group, which is the most popular architecture and widely used as a basic framework for advanced research by other researchers. The detail operations of each component in WZ coding process are explained by taking this solution as an example. Section 2.3 presents the WZ video coding solution from Instituto Superior Técnico (IST) together with the comparison with STD solution. The improved parts such as side information interpolation etc. are purposely emphasised. Another different WZ video coding approach called PRISM is reviewed in Section 2.4. In Section 2.5, other relevant improvements proposed by other researchers for WZ video coding are reviewed. Particularly, these algorithms are divided into different categories for review according to the targets of the proposed algorithms. Some other techniques relevant to the proposals of this thesis are discussed in later chapters for review.

2.1 Theoretic Background

Distributed Source Coding (DSC) theory refers to the coding of two or more dependent random sequences in a “distributed” way. Each source is independently encoded and the bits streams are sent to a single decoder for joint decoding in which the statistical dependencies are exploited. Before introducing the theoretical information theory results relevant to DVC, a brief description of the traditional source compression paradigm is given from the information theory point of view. The traditional source compression schemes are to answer those questions as follows: If we consider compressing multiple statistically dependent sources, i.e. two sources, what is the minimum encoding rate R to recover both sources without errors at a joint decoder? From Shannon Information Theory regarding to the compression of an independent identically distributed (i.i.d.) finite alphabet random sequence source X , it is stated that the source X can be reconstructed without loss only when the rate $R(X)$ is greater than or equal to its entropy $H(X)$. Hence, we can illustrate the above question by analyzing three possible typical situations as follows:

In Fig.2.1, two statistically dependent i.i.d sources X and Y are independently encoded and independently decoded. In this situation, the answer is simple as Information Theorem suggests that lossless reconstruction for both sequences can be obtained if $R(X) \geq H(X)$ and $R(Y) \geq H(Y)$.

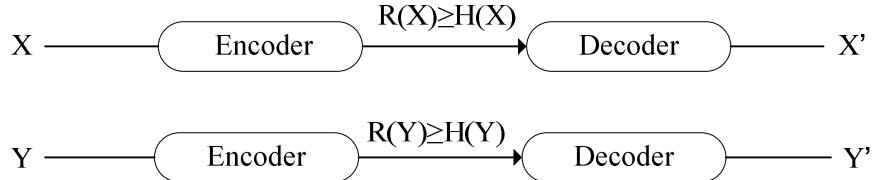


Fig.2.1 Architecture of independent encoding and independent decoding for two sources

In Fig.2.2, two statistically dependent sources X and Y are jointly encoded and jointly decoded. In this situation, Information Theorem suggests that a perfect reconstruction of X and Y can be obtained at decoder only when the rate R satisfies $R=R(X)+R(Y) \geq H(X,Y)$.

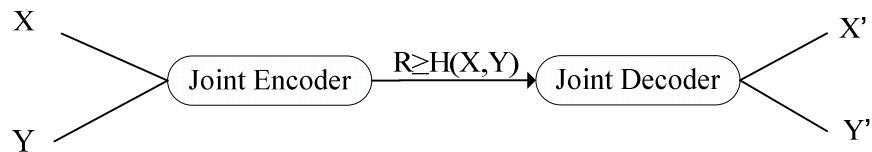


Fig.2.2 Architecture of joint encoding and joint decoding for two sources

In Fig.2.3 two statistically dependent sources X and Y are separately encoded but jointly decoded. This is the exact architecture of DSC. Clearly from traditional Information Theory, when R satisfies $R \geq H(X)+H(Y) \geq H(X,Y)$, X and Y can be accurately reconstructed at the decoder since they are correlated and the total bit rate R is higher than the joint entropy $H(X,Y)$.

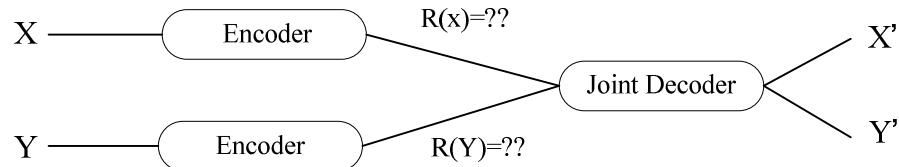


Fig.2.3 Architecture of independent encoding and joint decoding for two sources

2.1.1 Slepian-Wolf Theorem

Fig.2.3 illustrates the architecture of DSC. $R(X) \geq H(X)$ and $R(Y) \geq H(Y)$ are the answers from traditional Information Theory to this question. However, since the only necessary condition for the decoder to reconstruct these two sources without errors is that $R \geq H(X,Y)$ other than $R(X) \geq H(X)$ or $R(Y) \geq H(Y)$, if it is possible to achieve this lossless coding with $R(X) < H(X)$ or $R(Y) < H(Y)$? In fact, this problem has already been studied by Slepian and Wolf in [1]. Slepian and Wolf stated that: considering two statistically dependent i.i.d. sources X and Y , X and Y can be coded at rates $R(X) < H(X)$ and $R(Y) < H(Y)$ respectively and reconstructed with allowing an arbitrary small error probability after jointly decoding. However, there are several conditions needed to be satisfied:

$$R(X) \geq H(X|Y), R(Y) \geq H(Y|X), R(X) + R(Y) \geq H(X,Y) \quad (2.1)$$

where $H(X|Y)$ denotes the conditional entropy of X given Y and $H(Y|X)$ denotes the conditional entropy of Y given X . Equation (2.1) shows that even when the correlated sources are encoded independently, the total bit rate $R=R(X)+R(Y)$ equal to the joint entropy is enough just as that for joint encoding of X and Y .

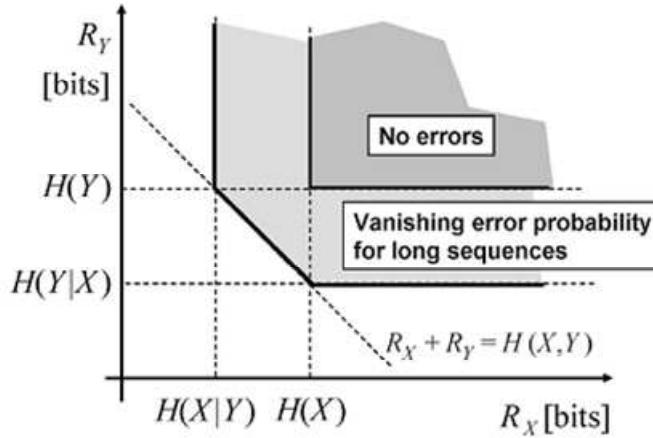


Fig.2. 4 Achievable rate region of Slepian-Wolf theorem [3]

According to the Slepian-Wolf theorem and equation (2.1), the rate region, also called Slepian-Wolf region, for the reconstruction allowing an arbitrary small error probability of X and Y can be described by Fig. 2.4. In Fig. 2.4, the vertical, horizontal and diagonal lines, corresponding to those three formulas of equation (2.1) respectively, represent the lower bounds for the achievable rate combinations of $R(X)$ and $R(Y)$. Slepian-Wolf coding generally refers to the lossless distributed source coding. Notice that lossless here is not mathematically lossless but allowing a controlled amount of errors which is approaching the lossless case.

One interesting feature of Slepian-Wolf coding is that it is a close kin to channel coding which already studied by Wyner [4]. Considering two i.i.d. binary sequences X and Y and a virtual “correlation channel”, the source sequence X and side information sequence Y are modelled as the input and output of the virtual channel respectively. Y is therefore a “noisy” version of X where “noise” introduced by the channel refers to the correlation between X and Y . Then a systematic channel code can be adopted to encode X and only the resulting parity bits are transmitted. At the decoder, the received parity bits and the side information Y are used to perform error-correcting decoding. In this approach, significant compression is resulted due to only few parity bits are needed to be sent.

2.1.2 Wyner-Ziv Theorem

Shortly after Slepian’s and Wolf’s work, Wyner and Ziv established information theoretic bounds

for lossy DSC compression with side information at the decoder [5]. Considering two i.i.d. random sequences X and Y (as shown in Fig.2.5) where X denotes the source sequence to be encoded at the encoder and Y denotes the side information which is only available at the decoder, the statistical dependency between X and Y is therefore unavailable at the encoder. A reconstruction of X , described by \hat{X} can be obtained at the decoder with the parity bits and side information Y . With a predefined threshold d , Wyner and Ziv established the minimum rate necessary to encode X guaranteeing \hat{X} with an average distortion below d . The results indicated that for the same threshold d , the minimum encoding rate (for X) of the case when the statistical dependency between X and Y is only available at the decoder, described by $R_{WZ}(d)$, is bigger than that of the case when the dependency is available both at the encoder and the decoder, described by $R_{X|Y}(d)$. The Wyner and Ziv theorem also can be described as

$$R_{wz}(d) \geq R_{X|Y}(d) \quad (2.2)$$

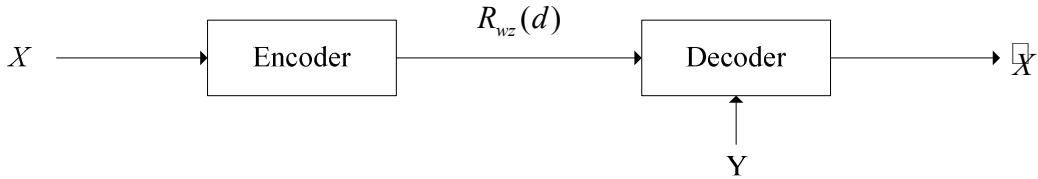


Fig.2.5 Wyner-Ziv theorem: lossy compression with side information

In the literature, $R_{WZ}(d)$ and $R_{X|Y}(d)$ are called Rate-Distortion (RD) functions. Notice that when $d=0$, equation (2.2) falls to the Slepian-Wolf result which means X is possible to be reconstructed with an arbitrarily small error probability even when the correlation between X and Y is only available at the decoder.

In conclusion, the Slepian-Wolf theorem and the Wyner-Ziv theorem work in a similar way, that is, compressing two statistically dependent sequences in a distributed way (separate encoding, jointly decoding) and the difference between them is that in the Slepian-Wolf theorem the dependency between two sequences is available both at the encoder and the decoder therefore sequences are lossless compressed with allowing an arbitrary small error probability between the source sequence and the reconstructed sequence, while in the Wyner-Ziv theorem the dependency is only available at the decoder and sequences are lossy compressed with allowing an predefined average distortion between the source sequence and the reconstructed sequence.

2.1.3 Practical Wyner-Ziv Code Design

It is known that Slepian-Wolf theorem is first proposed in the 1970s, however, it becomes popular only from the last few years since various applications emerged which motivated serious attempts at practical techniques. In 1999, Pradhan and Ramchandran [5] addressed the asymmetric case of

source coding in which the binary and Gaussian sources using scalar and trellis coset constructions are statistically dependent and the side information is available at the decoder. Later they considered the symmetric case [6-8] in which the source and side information are encoded with the same rate. In 2001, Wang and Orchard [6] made improvements over [5] by considering the asymmetric coding of Gaussian sources employing an embedded trellis code structure.

Since then, many channel coding techniques requiring iterative decoders have been adapted to the distributed coding problem. In 2001 and 2002, some researchers [7-9] proposed kinds of compression schemes in which statistically dependent binary sources are compressed by turbo codes. It has been proven that the turbo code based schemes are able to compress many kinds of sequences such as statistically dependent nonbinary symbols [10, 11], Gaussian sources [9, 12] as well as single sources [13, 14]. Meanwhile, Iterative channel codes have been investigated and proven that they can be used for joint source channel decoding if both the statistics of the source and the channel are available at the decoder [9, 13-15]. And the low density parity check (LDPC) codes have been suggested that might be a powerful alternative to turbo codes for distributed coding according to [16-19]. With turbo codes or LDPC codes, the compression performance can approach the Slepian–Wolf bound if the code performance approached the capacity of the correlation channel.

The work towards practical Wyner-Ziv coding schemes becomes active recently. In general, a Wyner-Ziv coder can be thought to be composed of a quantiser followed by a Slepian-Wolf encoder. For example, a Wyner-Ziv encoder can be implemented as a nested lattice quantiser [20, 21] or a trellis-coded quantiser [22, 23] followed by a Slepian-wolf encoder. The signal space is divided by the quantiser into cells which may consist of non-contiguous subcells. Fleming and Effros [23] presented rate-distortion optimised vector quantisers which either lack in performance or are prohibitively complex. Therefore, an efficient algorithm to find globally optimal quantisers among those quantisers with contiguous code cells was provided in [24].

With the development of practical Wyner-Ziv code design, the Wyner-Ziv coding theory is firstly applied into the video coding in [25], in which some frames of the video sequence treated as X source and encoded via Wyner-Ziv coding, which is implemented by a uniform quantised followed by a Slepian-Wolf encoder. Some other frames are coded via conventional intra-coding and almost perfectly reconstructed at the decoder. The decoder make use of the intra-coded frame and generate the estimation (side information, corresponding to the Y source) through frame interpolation. Then side information is used to decode the frame that coded in WZ way. In the next section, the WZ video coding which is main subject area of this thesis will be reviewed in detail.

2.2 Wyner-Ziv Video Coding-Stanford University Solution

In this section, popular WZ video coding architectures together with detail operations within each component from several leading research groups are firstly reviewed. Particularly, the WZ video coding from Stanford research group will be used as a main example to interpret the whole principle. WZ coding architectures from other research groups will be illustrated by comparing the differences with it from Stanford University.

2.2.1 Overall Architecture

Fig.2.6 shows a basic pixel domain Wyner-Ziv (PDWZ) video coding architecture from Stanford research group (STD) [3, 25], which is widely used as backbone for further development of many references.

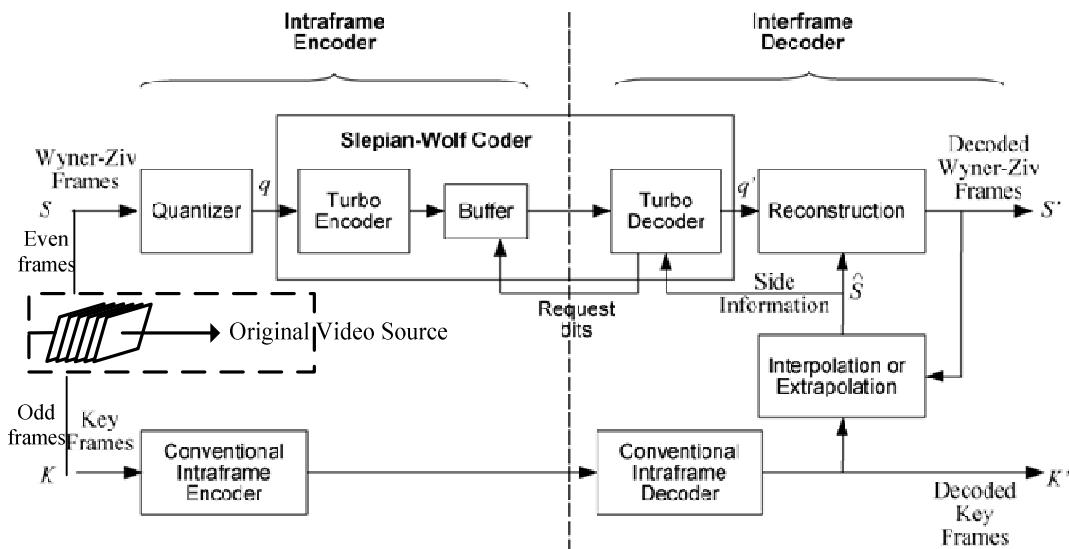


Fig.2. 6 STD pixel domain Wyner-Ziv video coding [3, 25]

In STD PDWZ, the frame sequence is categorised into two groups: WZ frame group composed of the even frames and key frame group composed of the odd frames. Suppose the temporal index of the video sequence is i , the WZ frame X_{2i} is intra-frame coded at the transmitter and inter-frame decoded at the receiver with the help of side information denoted by Y_{2i} . The key frames X_{2i-1} and X_{2i+1} are transmitted via conventional intra frame video coding, which are assumed to be reconstructed perfectly at the receiver in this architecture. The side information Y_{2i} is generated by interpolation between two of the received adjacent key frames. The side information is treated as noisy corrupted version of the coded WZ frame and used to decode the coded WZ frame at the decoder. More details about decoding process is introduced in Section 2.2.6&2.2.7.

- For WZ frame X_{2i} , each pixel is scanned row by row and uniformly quantised with $2^M = \{2, 4, 8, 16\ldots\}$ levels.

- The quantised symbols are binaried and fed into Slepian-Wolf codec for coding. Here,

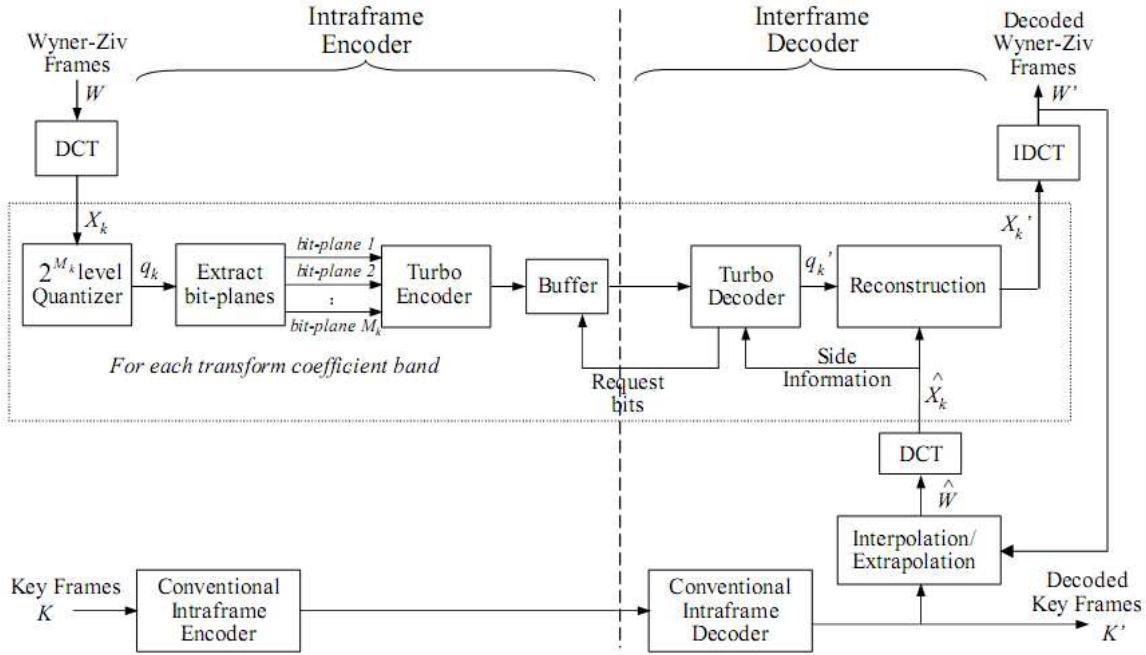


Fig.2. 7 STD transform domain Wyner-Ziv Video coding [26]

turbo codec is implemented as the Slepian-Wolf codec with the rate compatible puncturing method [27] to achieve flexible code rate to adapt the statistical change of side information Y_{2i} .

- After encoding, the parity bits are saved into a buffer for transmission and all the systematic bits are discarded.
- At the receiver, Laplacian distribution is used to model the statistic dependency between the original WZ frame X_{2i} and the side information Y_{2i} . With Y_{2i} , received parity bits and derived Laplacian distribution parameters, the turbo decoding can be operated followed by the reconstruction of quantised symbols.
- During the decoding, a “request-and-decode” process is adopted to realise the rate control. The decoder will initially request a fraction of parity bits from the encoder buffer to commence the decoding. If decoding fails, a request for more parity bits will be sent back to the encoder. An ideal error detection is assumed available at the decoder to judge if the decoding is successful or not. If the symbol error exceeds certain threshold, e.g. 10^{-3} , the decoding is regarded as unsuccessful. This process is repeated until the symbol error achieves the bound.
- The reconstruction of a complete WZ frame is performed with the reconstructed quantised symbols and side information Y_{2i} using $E(\hat{X} | q', Y_{2i})$. The whole coding process is finished

until all WZ frames are reconstructed.

The complexity of PDWZ is lowest comparing to that of the transform domain WZ video coding (will be shown below). It only requires a quantiser and a channel encoder. No ME or DCT transformation is required. However, in term of compression performance, it can be improved by introducing the DCT transformation with the expense of little added complexity.

Fig.2.7 shows the transform domain WZ coding (TDWZ) of STD group [26]. The operation is similar as that of STD PDWZ. The DCT transformation and bit-plane transmission are introduced in TDWZ. The system starts similarly as STD PDWZ by dividing the frame sequences into WZ frame group and key frame group.

- The WZ frame X_{2i} is DCT transformed by applying 4x4 blockwise DCT.
- Then the coefficients from the same position of each DCT block are picked out to compose the coefficient band.
- The coefficient band X_k is uniformly quantised and the quantised symbol stream q' is obtained.
- The bit-plane extraction is performed afterwards and the bit-plane will be sent to RCPT one by one according to the significance of the bit-plane, that is, from the most significant bit-plane to the least significant bit-plane.
- The generated parity bits are saved in the buffer and transmission is based on the request from the decoder. Meanwhile with this process, the key frames are transmitted to decoder via conventional intra-frame coding.
- The side information is interpolated between two adjacent key frames X_{2i-1} and X_{2i+1} . The same blockwise DCT as performed on WZ frame is applied to the side information Y_{2i} , and the coefficient bands are grouped by the same way as done at the transmitter.
- The correlation between the corresponding coefficient bands of X_{2i} and Y_{2i} is modelled as Laplacian distribution. With the coefficient band of side information Y_{2i} and received parity bits, current bit-plane is decoded, during which the “request-and decode” process is applied.
- After all bit-planes are decoded and quantised symbol stream q' can be reconstructed, following which the coefficient band is reconstructed by using $E(\hat{X}|q', Y_{2i})$.
- The WZ frame can be rebuilt after all coefficient bands are available and inverse DCT is carried out.

In [26], the performance of STD-TDWZ is evaluated. It is proved that STD-TDWZ offers superior performance to the STD-PDWZ due to the spatial redundancy exploited by DCT.

2.2.2 Quantisation

Quantisation is a technique in source coding to compress a range of values into a single value. The signal is divided into intervals or quantisation bins, which is labelled by a quantisation symbol. Since the number of bins is lower than the total number of the values that a signal can assume, a smaller bit stream is used to represent the signal and compression is achieved. In both PDWZ and TDWZ architectures, uniform scalar quantisation is utilised.

In order to get different rate-distortion (RD) points, $2^M = \{2, 4, 8, 16, 32, 64, 128, 256\}$ levels quantisation is applied in STD PDWZ. In the case of TDWZ, 7 examples of designed quantisation tables are used to quantise the DCT block shown in Fig.2.8, in which quantisation levels for the coefficient bands in the DCT block are defined.

<table border="1"> <tr><td>64</td><td>32</td><td>16</td><td>8</td></tr> <tr><td>32</td><td>16</td><td>8</td><td>4</td></tr> <tr><td>16</td><td>8</td><td>4</td><td>4</td></tr> <tr><td>8</td><td>4</td><td>4</td><td>0</td></tr> </table>	64	32	16	8	32	16	8	4	16	8	4	4	8	4	4	0	<table border="1"> <tr><td>64</td><td>16</td><td>8</td><td>8</td></tr> <tr><td>16</td><td>8</td><td>8</td><td>4</td></tr> <tr><td>8</td><td>8</td><td>4</td><td>4</td></tr> <tr><td>8</td><td>4</td><td>4</td><td>0</td></tr> </table>	64	16	8	8	16	8	8	4	8	8	4	4	8	4	4	0	<table border="1"> <tr><td>32</td><td>16</td><td>8</td><td>4</td></tr> <tr><td>16</td><td>8</td><td>4</td><td>4</td></tr> <tr><td>8</td><td>4</td><td>4</td><td>0</td></tr> <tr><td>4</td><td>4</td><td>0</td><td>0</td></tr> </table>	32	16	8	4	16	8	4	4	8	4	4	0	4	4	0	0	<table border="1"> <tr><td>32</td><td>16</td><td>8</td><td>4</td></tr> <tr><td>16</td><td>8</td><td>4</td><td>0</td></tr> <tr><td>8</td><td>4</td><td>0</td><td>0</td></tr> <tr><td>4</td><td>0</td><td>0</td><td>0</td></tr> </table>	32	16	8	4	16	8	4	0	8	4	0	0	4	0	0	0
64	32	16	8																																																																
32	16	8	4																																																																
16	8	4	4																																																																
8	4	4	0																																																																
64	16	8	8																																																																
16	8	8	4																																																																
8	8	4	4																																																																
8	4	4	0																																																																
32	16	8	4																																																																
16	8	4	4																																																																
8	4	4	0																																																																
4	4	0	0																																																																
32	16	8	4																																																																
16	8	4	0																																																																
8	4	0	0																																																																
4	0	0	0																																																																
\overline{M}^1	\overline{M}^2	\overline{M}^3	\overline{M}^4																																																																
<table border="1"> <tr><td>32</td><td>8</td><td>4</td><td>0</td></tr> <tr><td>8</td><td>4</td><td>0</td><td>0</td></tr> <tr><td>4</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td></tr> </table>	32	8	4	0	8	4	0	0	4	0	0	0	0	0	0	0	<table border="1"> <tr><td>32</td><td>8</td><td>0</td><td>0</td></tr> <tr><td>8</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td></tr> </table>	32	8	0	0	8	0	0	0	0	0	0	0	0	0	0	0	<table border="1"> <tr><td>16</td><td>8</td><td>0</td><td>0</td></tr> <tr><td>8</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td></tr> </table>	16	8	0	0	8	0	0	0	0	0	0	0	0	0	0	0	<table border="1"> <tr><td>16</td><td>8</td><td>0</td><td>0</td></tr> <tr><td>8</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td></tr> </table>	16	8	0	0	8	0	0	0	0	0	0	0	0	0	0	0
32	8	4	0																																																																
8	4	0	0																																																																
4	0	0	0																																																																
0	0	0	0																																																																
32	8	0	0																																																																
8	0	0	0																																																																
0	0	0	0																																																																
0	0	0	0																																																																
16	8	0	0																																																																
8	0	0	0																																																																
0	0	0	0																																																																
0	0	0	0																																																																
16	8	0	0																																																																
8	0	0	0																																																																
0	0	0	0																																																																
0	0	0	0																																																																
\overline{M}^5	\overline{M}^6	\overline{M}^7																																																																	

Fig.2.8 Quantisation matrices used for transform domain WZ coding [26]

2.2.3 Transformation

In image and video coding, the transformation is applied over $n \times n$ sample blocks of a frame to remove spatial redundancy. Since the neighbouring pixels within a block are typically correlated within a block, it is more efficiently to represent these pixels in the frequency domain. After the transformation, the block energy is mainly concentrated on few low frequency coefficients which include the DC coefficient and a small number of coefficients near to the DC coefficient. The compression can be achieved by only encoding these coefficients and neglecting the rest coefficients whose value are near to zero.

Similarly, in WZ video coding, transformations also can be used to achieve higher compression with the cost of little complexity added to the encoder. Due to the energy constrain, the entropy coding is not feasible in WZ video coding. In the TDWZ of [26], a coefficient band grouping method is proposed to encode the coefficients. Fig. 2.9 illustrates the whole process.

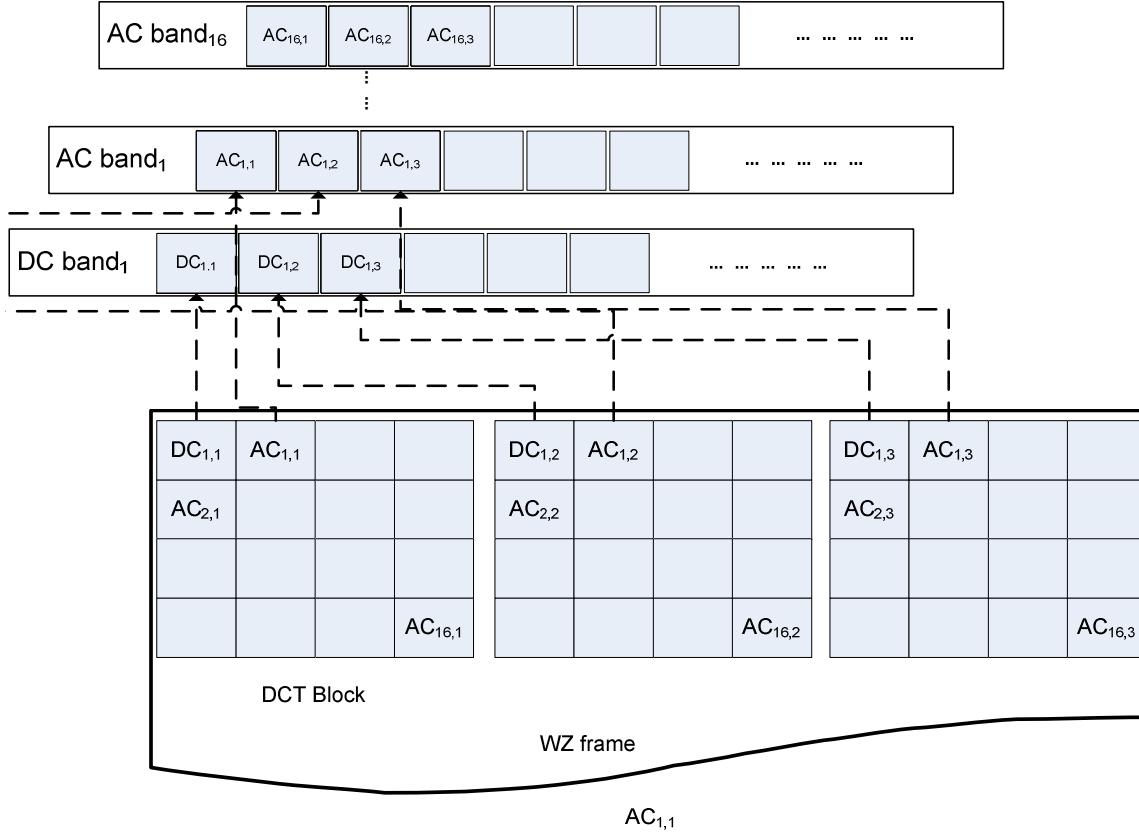


Fig.2. 9 Coefficient bands grouping in TDWZ

After the 4x4 blockwise DCT is applied to the WZ frame, coefficients located at same position in every DCT blocks are concatenated into one coefficient band. The way to organise the coefficients into bands has several following advantages:

- The correlation between WZ frame and side information can be more effectively exploited at the decoder. Since the correlation between corresponding coefficient bands of WZ frame and side information is high, the soft input of decoder can be more accurately predicted and less parity bits will be required during the decoding.
- For some AC coefficient bands, it can be directly replaced by same coefficient bands from side information. This operation leads to higher compression.
- It can be implemented in the applications such as unique protection scheme in error-resilience etc, in which certain coefficient bands e.g. DC coefficient band needs to be specially protected and different protection scheme can be applied.

2.2.4 Slepian-Wolf Encoder

The WZ codec is constructed using a quantiser followed by a Slepian-Wolf coder. As introduced in Section 2.1.3, the Slepian-Wolf codec can be implemented by a systematic channel codec, such as the turbo code applied in STD TDWZ architecture. Turbo code encoder is built using two

identical recursive systematic convolutional (RSC) codes with parallel concatenation as shown in Fig.2.10.

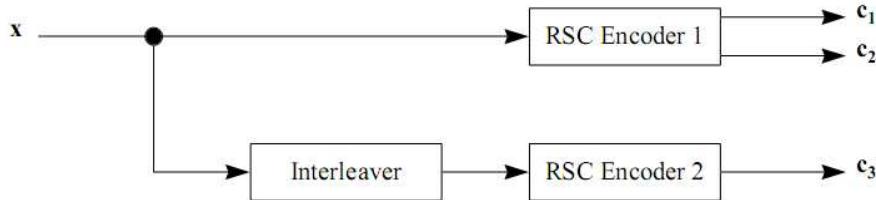


Fig.2. 10 The encoder architecture of turbo encoder

These two component encoders are separated by an interleaver. For each component encoder, it produce one systematic output streams and one parity output stream. Only the systematic outputs from the first component encoders is used, since the systematic output from the other component encoder is just an interleaved version of that output from the first encoder.

With turbo encoder, the code rate is usually 1/3. With the different puncturing pattern, various code rates can be achieved. The rate r is determined by the puncturing period p with $r=1/(p+n)$ where $n=1,2,3,..2p$. Fig.2.11 shows an example of puncturing pattern for code rate 1/2 with puncturing period of 8 for turbo coder, in which “1” means the bit at this position will be sent and “0” means the bit at this position is discarded.

1	1	1	1	1	1	1	1
0	1	0	1	0	1	0	1
1	0	1	0	1	0	1	0

Fig.2. 11 Example puncturing pattern for turbo encoding

After the puncturing, the encoder will concatenate the bits left and compose the output bit stream for transmission.

2.2.5 Feedback Channel

Since the encoder doesn't have any information of the side information, the feedback channel is widely used in WZ coding architecture to realise rate control during the coding process. The feedback channel is utilised by the so called “request and decode” process [3, 25]. At the beginning of transmission, all the generated parity bits are saved in a buffer and wait for transmission. Since the correlation exists between the side information and the WZ frame, the decoder first tries to decode just using the side information without any parity bits. If the side

information is extremely similar as the WZ frame, a successful decoding is possible. However, if decoding failed, the request for parity bits will be sent back to the encoder via the feedback channel. Based on the request, the encoder chooses the highest code rate determined by puncturing pattern and send corresponding number of parity bits. If decoding still failed, the encoder is requested to send another group of parity bits, which makes the total number of parity bits sent equals to the number of parity bits produced by next lower code rate. For instance, if puncturing period 8 is used, decoder should send the first subnet of parity bits which is produced by puncturing with code rate 8/9. If more parity bits are needed, the encoder sends additional parity bits thus the total number of parity bits sent equals to the number of parity bits produced by puncturing according to rate 8/10.

2.2.6 Side Information

The side information plays a key role in the WZ coding architecture. Side information is treated as a “corrupted” version of the WZ frame passing through the ‘virtual correlation channel’. In WZ coding, the correlation between the side information and the WZ frame is exploited at the decoder. Since the encoder has no knowledge of the side information during encoding process, the accuracy of the side information is consequently extremely important to the compression performance of WZ coding. If the side information is accurate enough, namely if it is very similar as the WZ frame, then few parity bits are required to be sent during decoding thus the compression is very efficient. Otherwise, the encoder has to send more parity bits to correct the “errors” between the side information and the WZ frame and the compression effect is not efficient obviously.

Therefore, it comes to the question how to generate accurate side information? In STD WZ coding architectures, only two adjacent frames are available (for the case of GOP=2). One needs to generate the side information from these two adjacent key frames. There are several ways to generate the side information in STD WZ architecture.

- i) Using one key frame as the side information.
- ii) Average interpolation. The side information is obtained by averaging the pixel value at same position in two key frames.
- iii) Motion compensated interpolation (MCI). The side information with time index t is generated by performing MCI between the decoded key frames at time $t-1$ and $t+1$. This interpolation technique involves symmetrical bidirectional block matching, smoothness constraints for the estimated motion and overlapped block motion compensation. Since the next key frame is needed for interpolation, the frames have to be decoded out-of-order,

similar to the decoding of B frames in predictive video coding.

- iv) Motion compensated extrapolation (MCE). To generate the side information for the even frame at time t , the motion between the decoded WZ frame at time $t-2$ and the decoded key frame at time $t-1$ is estimated using block matching and a smoothness constraint. The estimated motion is extrapolated to time t and the side information is formed by performing overlapped motion compensation using the pixel values from the previous key frame. Since a decoded WZ frame is used for ME, reconstruction errors from the WZ frame can degrade the reliability of the motion compensation. However, unlike MCI, the frames can be decoded sequentially.

The essence of side information generation is frame interpolation between two frames. There have been already numerous researches regarding to this method in the applications of frame rate up conversion, where the frame is skipped to be transmitted at the encoder and reconstructed at the decoder via frame interpolation techniques. However, the frame interpolation is based on the assumption that smooth motion lies between key frames and objects motion obeys the linear translation model. While this assumption is not always true, the interpolation quality can be degraded seriously when dealing with high motion sequences. It is fact that no available method can give confident results if no information of WZ frame is available during the interpolation.

2.2.7 Slepian-Wolf Decoder

The construction of a turbo decoder constituted by two identical Soft-Input Soft-Output (SISO) decoders is illustrated as Fig.2.12, the turbo decoder is called an iterative turbo decoder due to these two SISO decoders exchange information from each other.

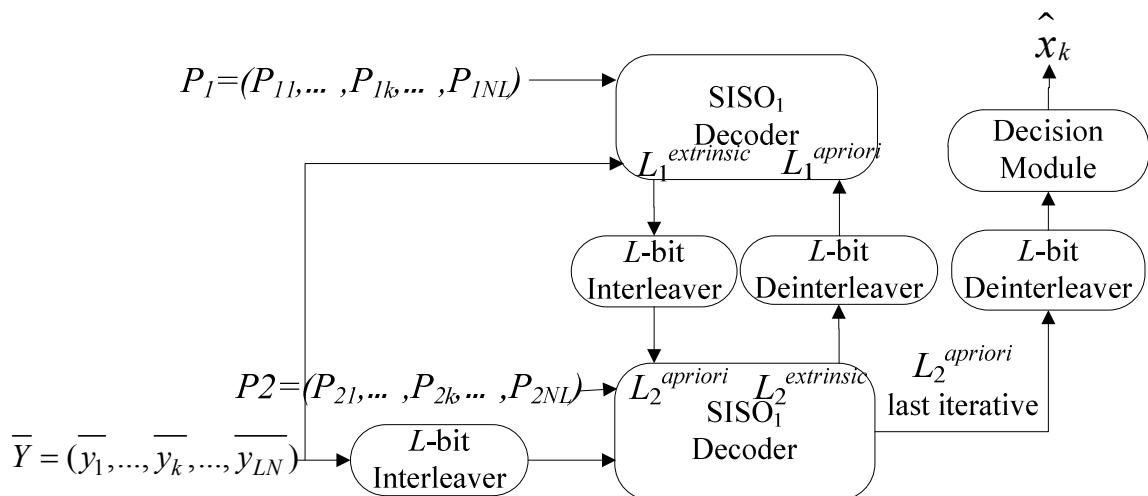


Fig.2. 12 SISO decoding in Turbo decoder [28]

Generally, the information inputting and outputting the SISO decoder is called soft input

information and soft output information, respectively. The term soft information means it cannot be directly determined which value it is instead it can only be determined which value it is more likely to be according to the probability.

As mentioned in above sections, the source systematic information X_{2i} is unavailable at the decoder since systematic information outputted by RSC encoder is discarded. Instead, a side information Y_{2i} , an estimation or a “noisy” version of X_{2i} is available at the decoder. Here “noise” refers to the differences between Y_{2i} and X_{2i} which can be expressed by r and modelled by Laplacian distribution as equation (2.3):

$$r = X_{2i} - Y_{2i}, f(r) = \frac{a}{2} e^{-a|r|}, a = \frac{\sqrt{2}}{\delta} \quad (2.3)$$

where δ^2 denotes the variance of r :

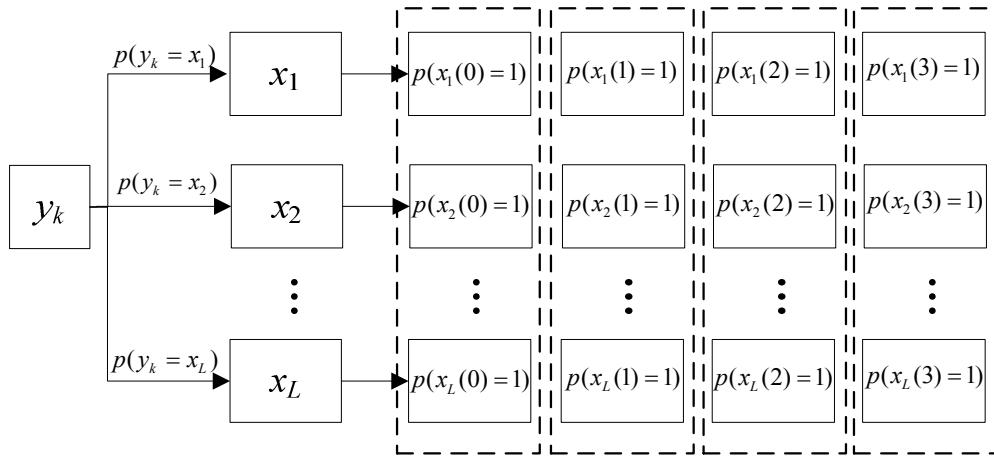


Fig.2.13 The Laplacian prediction process of soft input of the decoder

At the decoder, the received parity bits and the corresponding systematic bits of the side information are used to perform channel decoding. What we should note is that the systematic bits of side information cannot be directly obtained by quantising the pixels in Y_{2i} , instead they should be determined by considering the pixels of side information together with the correlation between Y_{2i} and X_{2i} , that is, the Laplacian distribution. It can be described in detail as follows:

- It is known that the residual between X_{2i} and Y_{2i} is modelled by Laplacian distribution as equation (2.3), according to which the estimation of bit streams can be obtained. Firstly we need to get the parameter α by calculating the variance parameter of the residual between X_{2i} and Y_{2i} . Then we can calculate the probabilities of each pixel in Y_{2i} according to equation (2.3). More precisely, a range for values of pixels in X_{2i} such as $[0, 255]$ is predefined or precalculated. The probabilities of a pixel y_k in Y_{2i} equals to each pixel in X_{2i} can be computed by following expression:

$$p(\hat{x} = x_m | y_k) = \frac{a}{2} e^{-a|y_k - x_m|} \quad x_m \in X_{2i} \quad (2.4)$$

Notice that y_k and x_m are pixels and \hat{x} denotes the prediction of input symbol.

- Next the pixel y_k is need to be quantised according to the probabilities obtained. For each pixel in X_{2i} , without loss of generality, it is assumed that the quantisation level is 16 which results in 4 bits per pixel and y_k is taken as an example and the process is illustrated as Fig.2.13.
- As shown in Fig.2.13, firstly the pixel x_m in the predefined region is quantised as $x_m(q)$ (where $m=1,2,\dots,L$ and $q=0,1,2,3$) where L denotes the number of pixels in the region. According to $p(y_k=x_m)$ we can obtain the probability of each quantised bit as following way:

- if $x_m(q)=1$ then $p(x_m(q)=1) = p(y_k=x_m)$ and $p(x_m(q)=0) = 1 - p(y_k=x_m)$,
- if $x_m(q)=0$ then $p(x_m(q)=0) = p(y_k=x_m)$ and $p(x_m(q)=1) = 1 - p(y_k=x_m)$.
- Then we have

$$p(\hat{y}_k(q)=1) = p(x_1(q)=1) + p(x_2(q)=1) + \dots + p(x_L(q)=1), \quad q = 0,1,2,3$$

$$p(\hat{y}_k(q)=0) = 1 - p(\hat{y}_k(q)=1)$$

where $\hat{y}_k(q)$ denotes the estimated value of the q th quantised bit of y_k .

- Finally, the input systematic bit $\tilde{y}_k(q)$ is determined by the Log-likelihood:

$$\tilde{y}_k(q) = L(\hat{y}_k(q)) = \ln \left(\frac{p(\hat{y}_k(q)=1)}{p(\hat{y}_k(q)=0)} \right) \quad (2.5)$$

where the operator $\ln()$ is the natural logarithm.

- Assuming that the input systematic bits from side information are denoted by $\bar{Y} = (\bar{y}_1, \dots, \bar{y}_k, \dots, \bar{y}_{LN})$. The parity bits generated by encoder are sent through the channel to the decoder regarded as channel information represented by P_1 and P_2 . With the channel information and the input systematic bits, we can go on with the turbo decoder. In the iterative turbo decoder depicted in Fig.2.12, firstly the soft output of the $SISO_i$ ($i=1,2$) decoder which is called extrinsic or reliability information is obtained and represented by $L_i^{extrinsic}$ ($i=1,2$). After that, $L_i^{extrinsic}$ is sent to the other $SISO_i$ ($i=2,1$) decoder as a priori information. The extrinsic information is exchanged between those two SISO decoders in order to improve the priori information of the other SISO decoder. Therefore, an iterative cooperation between two SISO decoders is realised by sharing the extrinsic information between them. Notice that the priori information of the $SISO_1$ decoder of the first iteration

is pre-established because the $L_2^{\text{extrinsic}}$ is not available yet.

- The decoding procedure may be performed using Maximum Likelihood (ML) algorithms or the Maximum A Posteriori (MAP) algorithm [29]. In 1974, Bahl and Cocke etc. proposed a modified algorithm of the MAP decoding. They showed that the MAP algorithm minimises the bit error rate when utilised in the decoding process of linear block and convolutional codes. In 1993, Berrou, Glavieux and Thitimajshima [30] employed the MAP algorithm to the iterative decoding of turbo codes. In our work, MAP algorithm is adopted.

2.2.8 Frame Reconstruction

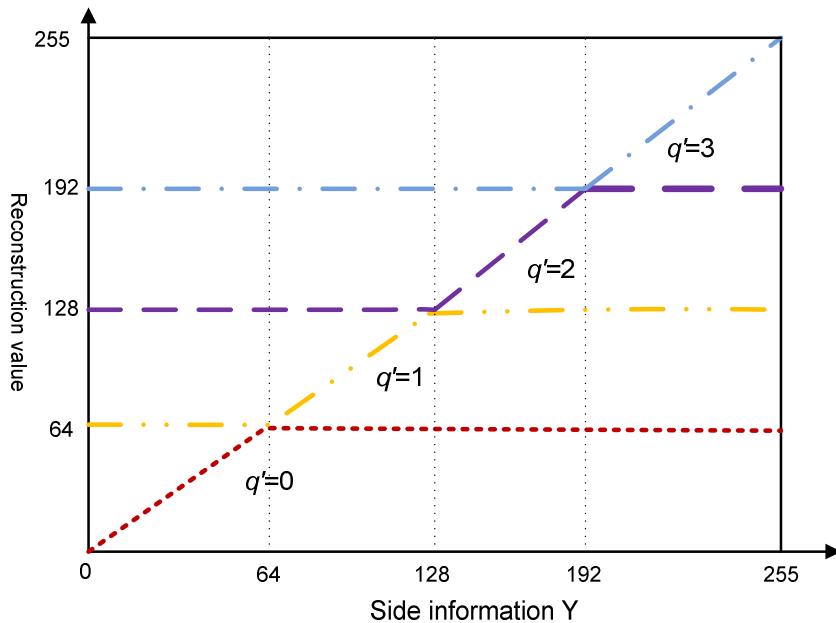


Fig.2. 14 Frame reconstruction function (4 quantisation levels) [25]

After decoding, the quantised symbol stream can be rebuilt as q' . The reconstruction of the whole WZ frame (or a coefficient band in TDWZ video coding) is performed by calculating the minimum mean squared error function $E(X_{\text{res}} | q', Y)$, where X_{res} denotes the reconstructed frame (for PDWZ video coding) or a coefficient band (for TDWZ video coding) and Y denotes the side information (for PDWZ video coding) or corresponding coefficient band of side information (for TDWZ video coding). If the side information lies within the reconstructed bin, the reconstructed pixel will take same value as the side information value. However, if the side information exceeds the quantisation bin, i.e., is outside the quantisation bin, the reconstruction function forces the reconstruction value to lie within the bin and takes the value of boundary of the bin.

The reconstruction function can be illustrated by Fig. 2.14 and alternatively , it can be

expressed by equation (2.6)

$$X_{res}(x,y) = \begin{cases} 2^{8-M} q'(x,y), & \text{if } q'(x,y) > Q'(x,y) \\ Y_{2i}(x,y), & \text{if } q'(x,y) = Q'(x,y) \\ 2^{8-M} (q'(x,y) + 1), & \text{if } q'(x,y) < Q'(x,y) \end{cases} \quad (2.6)$$

where (x,y) denotes the location of a pixel or coefficient, $q'(x,y)$ denotes decoded quantisation bin and $Q'(x,y)$ is the quantisation bin for the pixel or coefficient located at (x,y) in the side information Y_{2i} with same quantisation precision (2^M levels). If decoded quantisation bin is the same as the quantisation bin of side information, then $X_{res}(x,y)$ will take the value $Y_{2i}(x,y)$. If $Y_{2i}(x,y)$ is outside the decoded quantisation bin, then $X_{res}(x,y)$ will take the value of the bin boundary close to $Y_{2i}(x,y)$. This kind of reconstruction method limits the magnitude of the reconstruction distortion to a certain range determined by coarse level of quantisation.

2.2.9 Experimental Results

Stanford DVC research group mainly presents their research results for WZ video coding in [3, 25, 26, 31, 32], which have been widely used as references to compare performance by later researchers. Fig. 2.15 and Fig. 2.16 illustrate the performance of major range of WZ video coding solutions in transform domain and pixel domain.

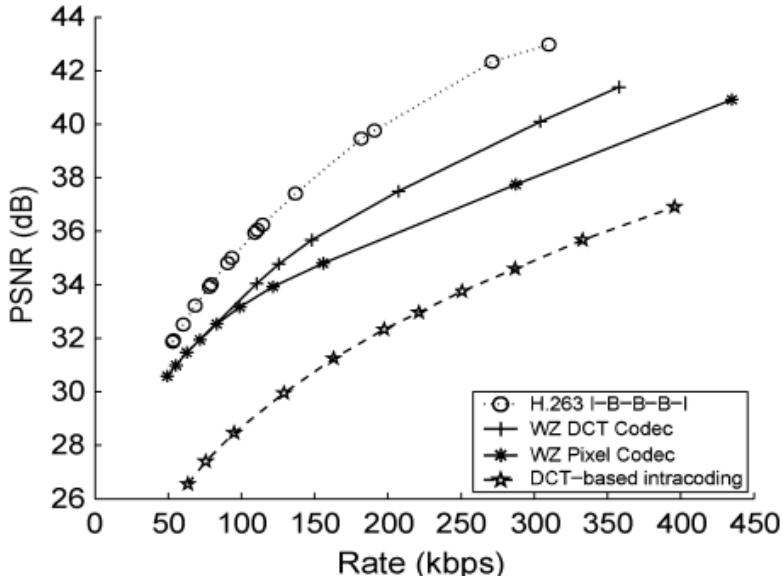


Fig.2. 15 RD performance of STD WZ codec with Salesman sequence [3]

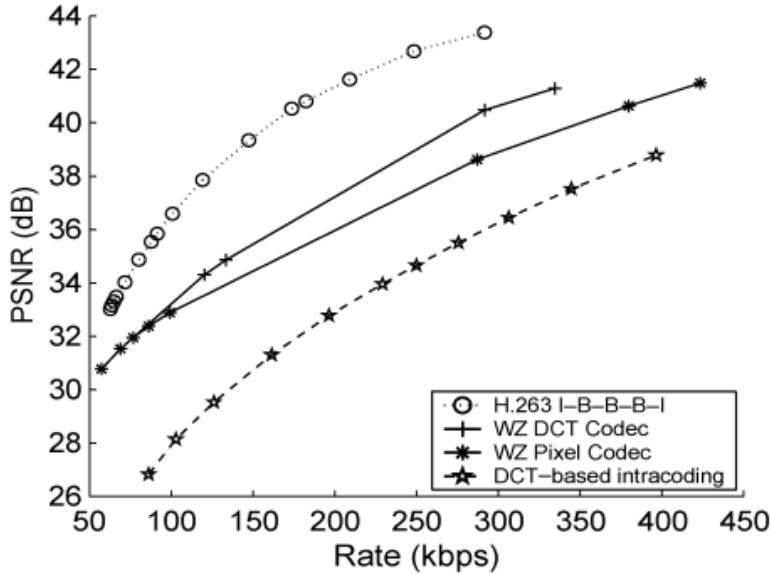


Fig.2.16 RD performance of STD WZ codec with Hall sequences [3]

In Fig.2.15 and Fig.2.16, the performances of STD TDWZ and STD PDWZ are compared with the traditional H.263 inter frame coding with IBIB structure and H.263 DCT intra frame coding. Fig.2.15 and Fig.2.16 show the PDWZ outperforms traditional DCT-based intra-frame coding up to 5 dB. TDWZ can further improve the performance and gain 2dB more than PDWZ. Nevertheless, both of them have significant gaps to the H.263 inter-frame coding. Note in the simulation, the GOP length is 4, which indicates that every fourth frame is the key frame and other frames are WZ frames. The length of GOP has direct influence to the overall system compression performance. However, longer GOP results in less complexity at encoder but inferior performance. In Arron's work [26], the algorithm regarding to how to design WZ video coding with different length of GOP is presented. Fig. 2.17 and Fig. 2.18 show the TDWZ codec with different length of GOP.

In Fig.2.17 and Fig.2.18, it shows that using larger size of GOP generally results in the total bit rate reduction. However, the enhancement in the compression performance with the increase in the size of GOP is not significant as it in H.263+ inter-frame coding. It can be observed in Fig.2.17 and Fig.2.18, TDWZ codec with GOP=8 performs similar as that with GOP=16 while in H.263+ inter-frame coding, GOP=16 outperforms than GOP=8 obviously. The reason is that in WZ coding, when the size of GOP is increased, there will be more WZ frames between two key frames. Less number of frames is considered as key frames thus the bit rate consumed for intra-frame coding is reduced. This is why the enhancement in overall compression performance is significant when GOP size increased from 4 to 8. However, if GOP size is too large, although the bit rate consumed on encoding the key frames are reduced, the average bit rate used to encode the

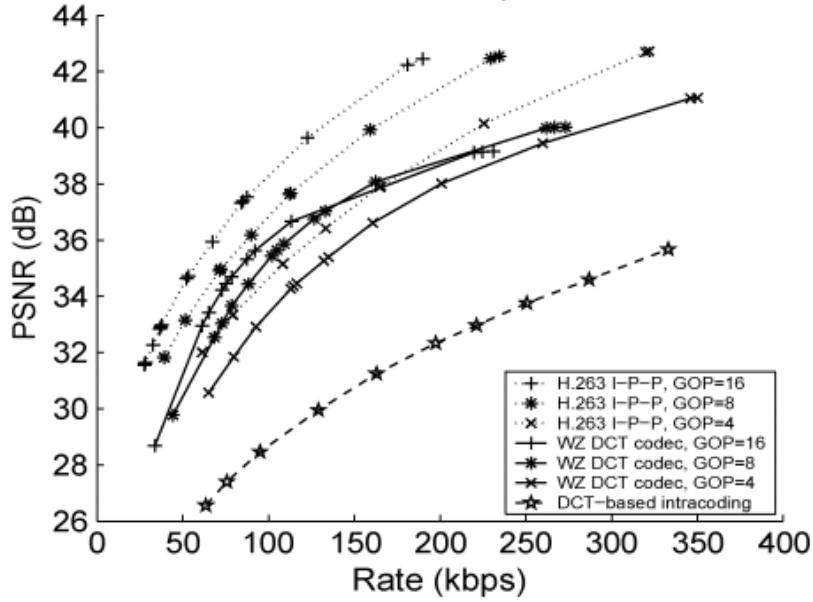


Fig.2. 17 RD performance of STD TDWZ with different length of GOP for Salesman [3]

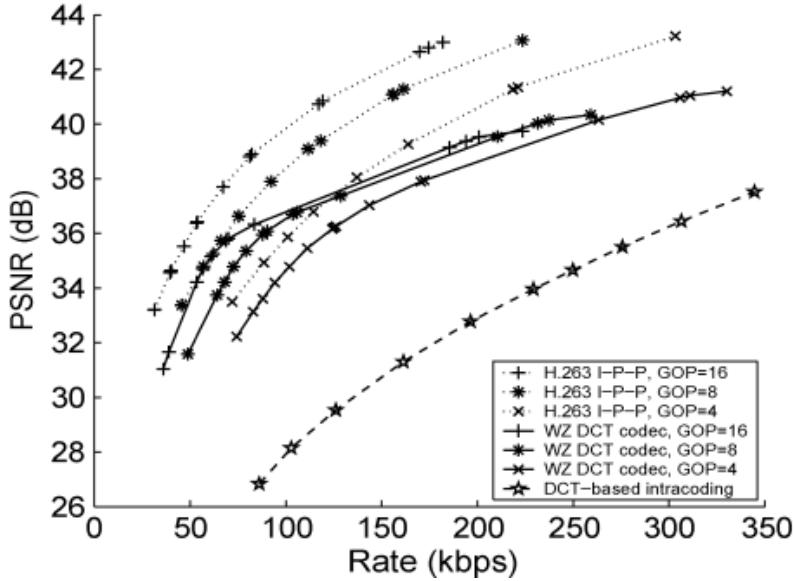


Fig.2. 18 RD performance of STD TDWZ with different length of GOP for Hall [3]

WZ frames can be increased. Larger size of GOP also denotes further distance between two key frames, which could easily result in the bad motion compensated interpolation, namely can produce the side information with much inferior quality. The low quality side information leads to the severed increase in the number of parity bits during decoding thus bring down the overall compression performance. Consequently, in WZ coding, the GOP size cannot be too large and highly depends on video type.

Although STD WZ is not the earliest practical WZ video coding solution, it has been largely

used as platform to build advanced WZ video coding in research. However, in the original publications of STD WZ solutions, some important details are missed.

2.3 Wyner-Ziv video coding- Instituto Superior Técnico Solution

In this section, some major research achievements from Instituto Superior Técnico (IST) are presented. IST is the one of leading research groups working for project “Distributed Coding for Video Services” (DISCOVER) which has made lots of research achievements around DVC area based on the architecture of STD WZ solution. In this section, several improvements proposed by IST research group are reviewed with comparison to that of the STD WZ video coding architecture.

2.3.1 IST PDWZ

In [33], the PDWZ video coding is proposed by Ascenso et al in 2005. The backbone of the architecture is based on the STD PDWZ video coding introduced in Section 2.2.1. However, there are two major different parts that are bit-plane transmission and side information generation.

2.3.1.1 Bit-plane Transmission of IST PDWZ

Bit-plane transmission is used in TDWZ video coding of STD solution for transmitting the coefficient bands. In fact it can be used in the PDWZ video coding as shown in Fig.2.19. Each pixel is scanned and quantised, after which the bit-plane extraction is performed by grouping the bits of quantised symbols with same significance into a bit-plane. The bit-plane will be sent to the Slepian-Wolf encoder subsequently from the most significant bit-plane to the least significant bit-plane. Only the previous bit-plane is successfully decoded at the decoder, next bit-plane can be transmitted. In [34], it also shows that the order of bit-plane transmission can be inverted by sending the least significant bit-plane first and most significant bit-plane finally. This change can speed up the decoding speed up to 30%. It is indicated by Arron in [25, 31] that the performance of PDWZ video coding with or without bit-plane transmission is similar. However, bit-plane extraction can reduce the size of the input block for Slepian-Wolf codec and have effects in reducing the decoding complexity in the Slepian-Wolf decoder.

In [33], the IST PDWZ video coding outperforms the STD PDWZ video coding. The enhancement of the performance of IST PDWZ video coding can largely attribute to the more complicated side information generation technique which is introduced in next section.

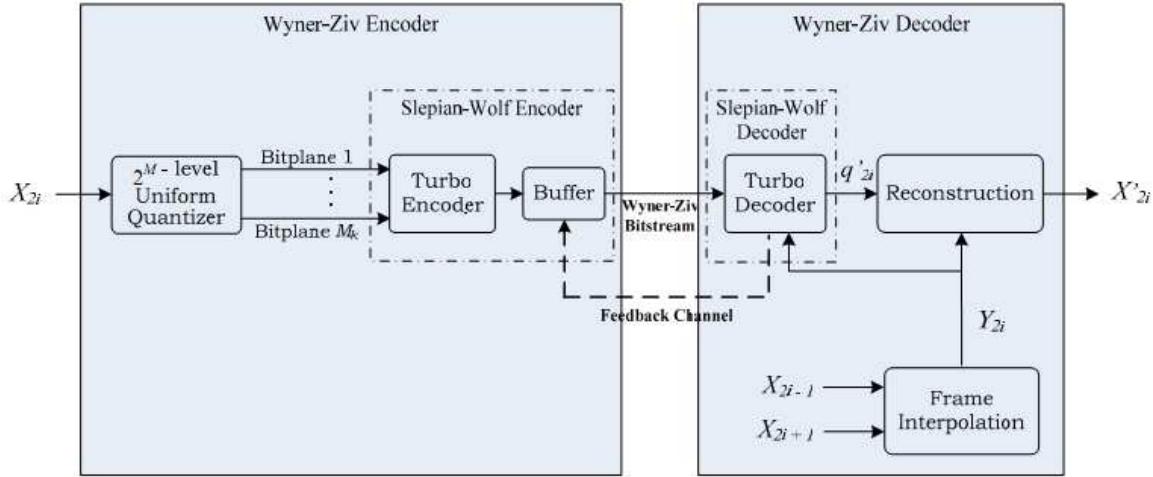


Fig.2. 19 The architecture of IST PDWZ [33]

2.3.1.2 Side information generation of IST PDWZ

In [33], a rather complicated side information generation method is proposed, which is later used in most IST references. The method works for interpolating side information between two adjacent key frames. Several process parts are included as shown in Fig. 2.20

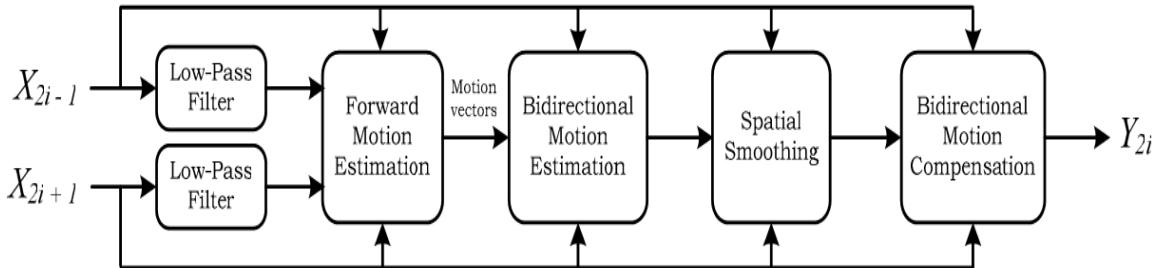


Fig.2. 20 Side information generation framework for IST PDWZ[33]

Low pass filter: The key frames are firstly passed through the low pass filter to remove noise and improve the accuracy of following ME.

Forward ME: The full block searching algorithm is used. The motion vectors obtained is used to compose a coarse motion vector field for the later interpolation. The details are described in Fig. 2.21.

Bidirectional ME: With the motion vectors obtained in previous step for the block to be interpolated, a further bidirectional motion search is performed to search two most matching blocks in two key frames respectively. The bidirectional motion vector is fixed after the bidirectional ME and can be used directly to interpolate the frame. The details are described in Fig. 2.22.

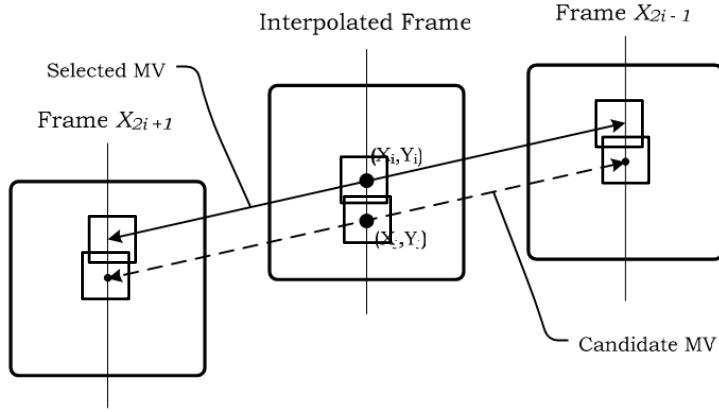


Fig.2. 21 Forward ME for side information interpolation in IST PDWZ[33]

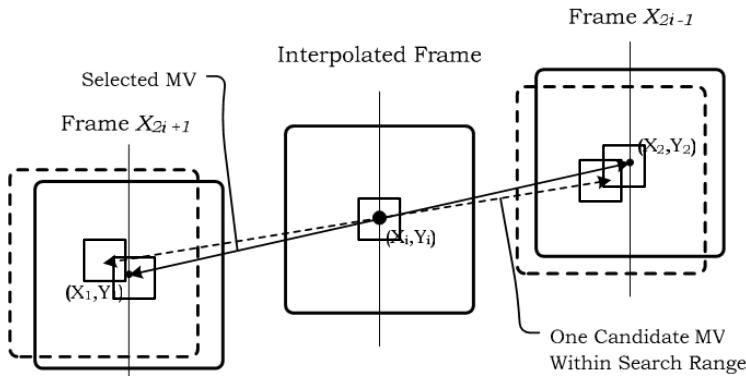


Fig.2. 22 Bidirectional ME for side information interpolation in IST PDWZ[33]



Fig.2. 23 Perceptual comparison for side information with or without spatial smoothing [33]

(Left: with spatial smoothing. Right: without spatial smoothing)

Spatial smoothing: Before interpolation, a spatial smoothing process is applied to correct the false motion vectors and improve the spatial coherence. The main algorithm used is a weighted vector median filter [35] which is extensively used for noise removal in multichannel images.

Bidirectional motion compensation: With the final bidirectional motion vectors, the side

information is interpolated by averaging two pointed block by the motion vectors. Fig.2.23 shows the perceptual comparison for side information interpolated by bidirectional motion compensated interpolation with or without spatial smoothing process.

2.3.2 IST TDWZ

The IST TDWZ is mainly presented in [28], the architecture is not shown here since it is same as that of STD TDWZ. However, the inside operations of IST TDWZ are slightly different from those of STD TDWZ, and the system performance of IST TDWZ outperforms that of STD TDWZ. Those operations are listed below:

Generation of side information: The side information generation method used in IST PDWZ is used here. No matter it is applied to TDWZ or PDWZ, better side information can offer better system performance.

Quantisation of DCT coefficients: In IST TDWZ, the value range of the AC coefficient bands is known at the decoder. With this operation, the quantisation loss can be minimised. If the same number of quantisation levels is applied to a value range which is smaller than a fixed value range, then a smaller quantisation interval is resulted which means that a smaller quantisation step size is resulted. Therefore a lower distortion of output can be obtained. However, this operation may result in a slight increase in the bit rate, which is recorded in the simulation as 240 bits per frame.

2.3.3 Experimental Results

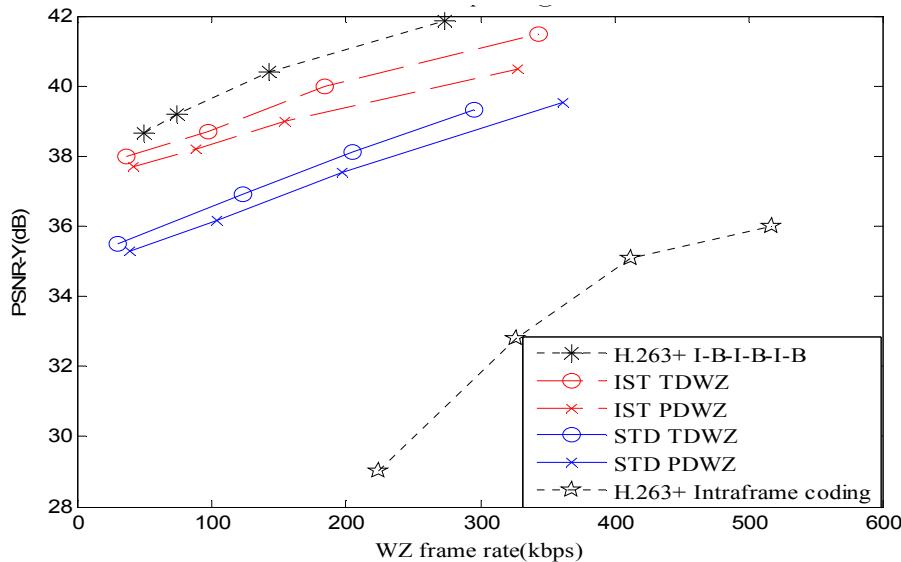


Fig.2. 24 PSNR comparison for IST and STD WZ codec with Foreman sequence@30HZ

In Fig.2.24, the results of STD TDWZ,STD PDWZ,IST TDWZ and IST PDWZ are presented together to make comparison. The results are from references [28, 31], in which the first 101

frames of Foreman sequence (QCIF) are simulated with frame rate of 30HZ.

In Fig.2.24, the RD plot clearly shows that the performances of WZ coding solutions are much superior to those of the traditional H.263 DCT intra frame coding. However, there is still a large gap between WZ coding solutions and the traditional H.263 inter-frame coding with IBIB structure. It is evident that for both research groups IST and STD, the TDWZ coding usually outperforms PDWZ coding. It is surprisingly found IST solutions outperform STD solutions due to more advanced side information generation algorithms are used in IST solutions. The IST PDWZ even outperforms STD TDWZ more than 2 dB.

In this section, we mainly introduced the basic structure of IST WZ video coding. This architecture later has been further improved and evolved into ‘DISCOVER’ codec. A details evaluation of DISCOVER codec can be found in [36]. However, we will not present the codec of DISCOVER here since several advances covered by DISCOVER codec are introduced specifically in section 2.5.

2.4 Wyner-Ziv Video Coding-PRISM Solution

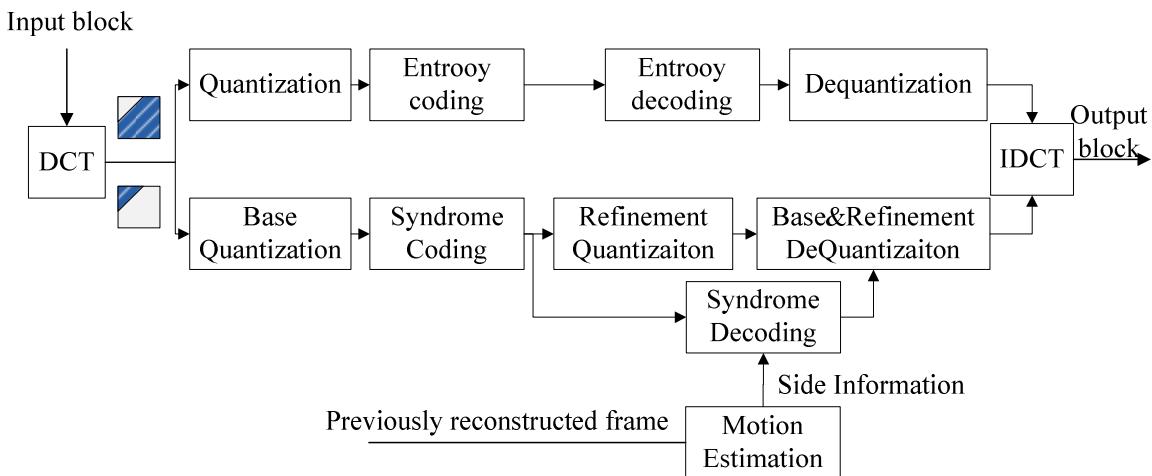


Fig.2. 25 The architecture of PRISM [37]

PRISM presented in [37-41] is a quite different WZ video coding approach comparing to the STD and IST solutions. PRISM refers to “Power-efficient, Robust, high-compression, Syndrome-based Multimedia coding”, which is proposed by Berkeley’s research group. The difference between PRISM and IST or STD WZ coding solutions is described as follows:

- i) Syndrome coding is used instead of turbo coding in STD and IST solutions. As indicated in [3], the essence of using syndrome coding is same as using a systematic channel codec, e.g. turbo codec.
- ii) The WZ coding is based on the block level. But in IST or STD solutions, the coding is based on frame level.

- iii) There are more than one side information available at the decoder during decoding the current block, while in IST and STD solution, only one side information is used to decoded one encoded WZ frame.

Fig. 2.25 shows the basic architecture of PRISM. In following sections, the encoding and decoding procedure are briefly introduced and the experimental results are discussed.

2.4.1 Encoding

The input video frame is divided into blocks e.g. 8x8 or 16x16. All of operations are based on the block level.

Classification: Blocks in current frame will be divided into 3 classes: skipped class, intra coding class and syndrome coding class. For skipped class, the block will be skipped for transmission. For intra coding class, traditional intra coding is applied. For syndrome coding class, the PRISM coding is used as shown in Fig.2.25. The criteria to judge the class type of the block is based on the correlation evaluation between the block and the block co-located in previous reconstructed frame. Particularly, the square error between blocks is used to model correlation intensity. After the classification, the information of the class of the block will be sent to the decoder as header information.

Transformation: For each syndrome coding class block, a blockwise DCT is applied, after which the coefficients located at top left corner represent the most energy of the block. The coefficients are Zig-Zag scanned and divided into two fractions for different quantisation.

Quantisation: The DC coefficients together with few AC coefficients nearby will be coarsely quantised, in which the quantisation step is determined by the correlation level defined in Classification stage. For the other AC coefficients (high frequency coefficients), traditional quantisation and entropy coding are applied. Since high frequency coefficients are small value or value near to zero, the traditional approach encodes the high frequency coefficients with small amount of bits. While for low frequency coefficients, the value usually is large and syndrome coding can encode the coefficients with fewer bits than normal coding since only syndrome coset index will be sent, therefore the size of quantised bits is much smaller than the original length.

Syndrome coding: In syndrome coding, the quantised codewords are grouped into several cosets. Each coset has an index label (syndrome) associated with. It is the syndrome other than original quantised symbol codeword sent to the decoder. Since the number of syndromes is smaller than the number of codeword, the compression is achieved. This operation is similar as applying systematic channel coder to code the quantisation symbol and only parity bits are sent in STD or

IST WZ coding solutions. In [39], a trellis-based syndrome code (128-state rate- $\frac{1}{2}$ trellis code) is used to quantise lower frequency coefficients. The syndrome bits are projected into the bit stream syntax at the block level.

Refinement Quantisation: In order to get a better reconstruction quality, a refinement process for the base quantisation step size is applied. The refinement process divides the base quantisation interval into many smaller sub-partitions, whose number is determined by the desired reconstruction quality. These sub-partitions are called refinement intervals which are assigned indexes individually. The refinement bits of the refinement interval index are transmitted to the decoder; these bits are another component of the bit stream syntax at the block level.

Cyclic Redundancy Check (CRC): CRC is applied to the quantised sequence before applying above procedures. The decoder searches all available predictors “Y” to decode the sequence labelled with syndrome bits. If CRC check is matched after the decoding with current predictor “Y”, the decoding is declared to be successful.

Syntax composition: The bit stream syntax for current block consists of following parts as shown in Fig.2.26.

Syndrome bits	CRC bits	Refinement bits	Pure source code bits
---------------	----------	-----------------	-----------------------

Fig.2. 26 Bit stream syntax in PRISM[37]

Note that the “Pure source coded bits” is the bits obtained after the quantisation and entropy coding of high frequency coefficients.

2.4.2 Decoding

For those blocks defined as skipped class, the block co-located in previous frame are used instead. For those blocks defined as intra coding blocks, the inverse operation is performed. The following procedure shows the decoding process for the blocks defined as syndrome coding.

Motion estimation (ME): ME is performed at the decoder with previous reconstructed frame. All neighbouring blocks within the searching range are used as candidate predictors, namely side information to decode the quantisation codeword with syndrome bits. During the ME process, the full search ME algorithm with half pixel accuracy is used.

Syndrome decoding: With received parity bits and predictor “Y”, a set of possible quantisation codeword sequences are chosen. In order to find the best result, the Viterbi algorithm is used. Then CRC check is performed to check the accuracy of the decoding. If CRC check is unsuccessful, the next predictor is used. The process is repeated until a successful decoding is obtained.

Base and Refinement Dequantisation: After reconstruction the quantised coefficients are reconstructed, and then the base dequantisation and the refinement dequantisation with the refinement bits transmitted from the encoder are performed. During the recovery of the decoded block coefficients, the predictor used for decoding is used to give the best reconstruction of the block.

Other procedures: Entropy decoding and dequantisation are performed to obtain the high frequency coefficients. With the decoded skipped class blocks, intracoding blocks and syndrome coding blocks, the inverse DCT is performed and the output frame is obtained.

2.4.3 Experimental Results

In the simulations of [38-41], some video paradigms with low complexity are proposed with efficient compression performance with comparison to that of H.263+. The PSNR performance of PRISM lies between that of H.263+ intra-frame coding and that of Inter-frame coding. However, the virtual quality is indistinguishable comparing to that of the H.263 inter-frame coding. Moreover, in the simulation for robust transmission, PRISM shows superior performance to the H.263+ coding due to the effective measures are used to stop the error propagation.

One of the major differences between PRISM and WZ video coding solutions proposed by STD and IST is that PRISM is operated on the block level, while the other two are based on frame level. Working on the block level may limit the PRISM application in multi cameras where frame level WZ coding is needed. However, the PRISM is superior in the aspect of robust transmission, which is not covered by WZ video coding solutions proposed by STD and IST.

2.5 Other Advances in Wyner-Ziv Video Coding

In this section, the other advances in WZ video coding from research community are reviewed. There are numerous approaches regarding to DVC have been researched in last few years. Due to the length limitation, we cannot review all of them but pick some references with significant progress towards an advanced DVC for review. Specially, these references were divided into several categories according to their applications emphasised in the references.

2.5.1 Advances in Quantisation

Uniform scalar quantisation is popularly used in most WZ coding solutions. However, more sophisticated quantisation techniques can be engaged to further improve the WZ coding performance. The only problem is if a sophisticated quantisation technique resulted in higher complexity at WZ encoder. In [42], the Lattice Vector Quantisation(LVQ) was employed based on a WZ coding solution with LDPC. The adopted LVQ offers better coding performance than scalar

quantisation but still with low complexity. The similar work also can be found in [43]. In [44, 45], two modified versions of Lloyd maximum quantisation are used to replace the uniform quantisation in DVC. Both algorithms are claimed to have advantages in reconstruction process and superior to the uniform quantisation. Moreover in [46], a nested scalar quantisation is used in their framework.

2.5.2 Advances in Transformation

In above WZ coding solutions, DCT is frequently used in TDWZ video coding. However, many researchers have investigated the applications of DWT in TDWZ coding. DWT inherits advantages in reducing the block artifacts and scalability over DCT. There have been several approaches for the wavelet based DVC, in which two main algorithms EZW [47] and SPIHT [48] are used to encode the wavelet coefficients. In [49], DVC based on DWT is proposed, in which the spatial scalability and SNR scalability are also explored. The proposed system is robust to the channel errors. Other works regarding to the scalability of wavelet based DVC can be found in [50, 51]. In [52], Zerotree entropy coding(ZTE) is used to encode the wavelet coefficients in DVC. The quantised wavelet coefficients are reorganised in terms of the zero tree structure to distinguish the significant coefficients and insignificant coefficients. The significance map is intra-coded and transmitted while the significant coefficients are Slepian-Wolf coded. In [53], DVC based on SPIHT is proposed where low frequency coefficients are coded using WZ codec and high frequency coefficients are coded using SPHIT. The algorithm has better performance than the intra coding algorithm which purely employing SPIHT. In [54], DVC based on JPEG 2000 framework is proposed. In this architecture, the key frames are coded using JPEG 2000 intra coding, while the WZ frame is processed by an optimum quantisation algorithm and Gray codec, with which scalability is also reached. In [55], An efficient DVC scheme combining residual coding, Slepian-Wolf SPIHT (SW-SPIHT) coding and intra-mode decision is proposed. Unlike references quoted above, the scheme adopts the wavelet domain WZ codec to compress the residual frame of the current frame and its reference frame. An intra-mode decision based on the temporal and spatial correlation is made to determine whether a wavelet block should be coded using intra-SPIHT coding or SW-SPIHT coding. Evident improvements comparing to available references are observed. There are still some other wavelet based WZ coding solutions which are not covered in this section. However, the wavelet based WZ coding generally shows higher performance than intra frame coding algorithm and PDWZ coding solutions.

2.5.3 Advances in Slepian-Wolf Codec

One of the alternative choices to the turbo codec used in current WZ coding is Low Density

Parity Check (LDPC) codec. LDPC codec is a new powerful systematic channel codec and it is proved in many references that LDPC has similar performance as or even superior performance to turbo codec. LDPC based Slepian-Wolf codec is widely adopted in [42, 46, 50, 54, 56-58] etc. Most of them work in the similar way as turbo codec used currently. The bit-plane extracted from quantisation symbols is sent to LDPC encoder, with which syndrome coding is operated and only the syndrome bits generated are sent to the decoder. The decoder is performed with the side information and syndrome bits. Since most references employ LDPC as basic Slepian-Wolf codec to propose other more advanced techniques, it is hard to make decision if LDPC is superior to the turbo codec for WZ coding solutions. In the work of[59], the symbol based approach and bit-plane based approach in WZ coding with LDPC are compared and the conclusion that both has similar performance is given. However, bit-plane based approach has advantage in reducing the computation thus is preferable in the application. In addition to LDPC codec, the digital fountain code for WZ coding is proposed in [60].

2.5.4 Advances in Modelling Correlation Channel

The modelling of the correlation channel in current WZ coding solutions is very important for the system coding performance. Without an appropriate modelling, the prediction of the soft-input information in Slepian-Wolf decoding tends to be wrong, thus the whole decoding performance will be changed to rely on the parity bits received, which caused inefficiency in decoding inevitably. In most WZ coding solutions, the correlation channel (or virtual channel, or correlation noise, or error residual distribution between WZ frame and side information for PDWZ case, or residual distribution between corresponding coefficient bands of WZ fame and side information for TDWZ coding case) is modelled by Laplacian distribution. In the WZ coding architecture, the Laplacian parameters should be estimated by using the side information and the WZ frame. However, since the WZ frame is not available at the decoder and side information is not available at the encoder, these parameters usually are estimated offline for researchers. In STD WZ coding solutions [3, 31, 32], it stated the Laplacian parameters can be estimated by using side information and one of the adjacent key frames, or one previously decoded WZ frame. This operation didn't degrade the system performance as proved. In Brites' work [61, 62] from IST, an online correlation noise model estimation algorithm is proposed at the decoder for both PDWZ coding and TDWZ coding. In this algorithm, estimation is performed based on different levels of granularity. Three levels of granularity which are frame, block, and pixel levels are used for PDWZ coding while DCT bands and coefficients are two granularity levels used for TDWZ coding. The higher granularity for estimation leads to the better rate-

distortion performance. It is found that the pixel levels and coefficients levels are the best solutions for PDWZ coding and TDWZ coding respectively. In [63, 64], both authors mentioned that the side information quality varies from point to point due to the video nature content e.g. occlusion area. Therefore modelling correlation noise as stationary statistics is not the best solution for the context of PDWZ. In both works, the algorithms of modelling correlation noise as non-stationary statistics are proposed, in which different Laplacian parameter is used for each pixel of the side information and evident enhancement in WZ coding performance can be observed in their simulations. Similar work can also be found in [65]. In [66], a probability model for DVC with multiple side information is considered. The transitional probability is calculated from the conditional probabilities on the multiple side information signals. In [67], works toward modelling correlation noise between wavelet sub bands are explored for wavelet based DVC solution.

2.5.5 Advances in Side Information Generation

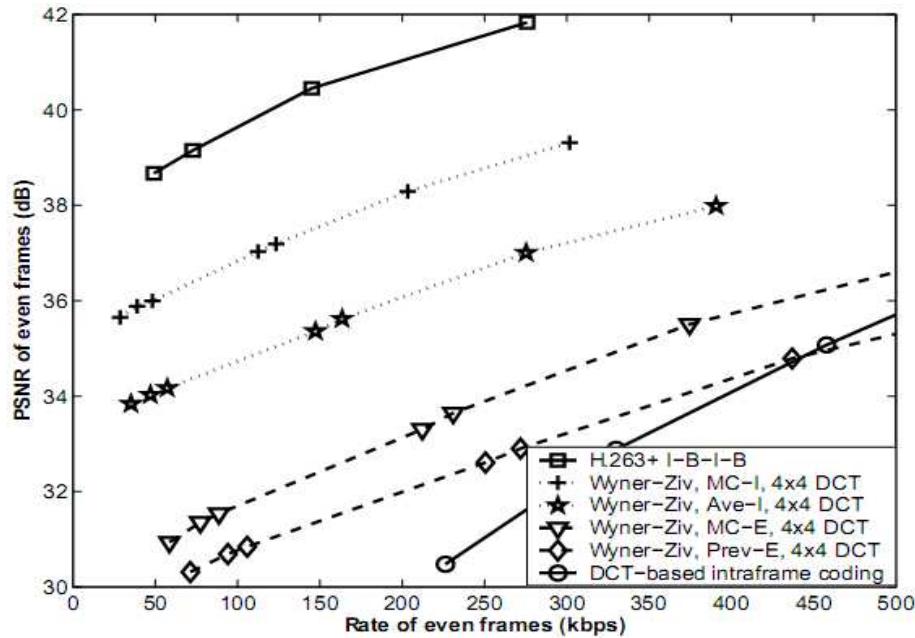


Fig.2. 27 RD performance of STD TDWZ with different side information interpolations for Foreman [31]

Side information is another key factor which can significantly influence the WZ coding performance. From Fig.2.27 of STD's work [31], it is evident that different quality of the side information results in different WZ coding performances eventually. It can be seen that between the simplest average interpolation and complicated MCI interpolation, the PSNR gap can be more than 6 dB. In Fig.2.27, the reason why IST's solution outperforms STD's solution can largely attribute to more complicated and advanced side information generation techniques used.

Besides the algorithms used in IST and STD solutions, there are many other approaches in the research community regarding to generating high quality side information. One of pioneering research is to refine the side information by using decoded WZ frame. Then the whole decoding process will repeat with the refined side information and the better output can be obtained. The side information can be refined again and this process can be repeat for several times until the quality of output is fixed. This type of work is mainly presented in [68, 69] which will be reviewed in more details in Chapter 3. In WZ architecture of [70], the encoder sends hash codeword of the current frame to aid the decoder to accurately estimate the motion, with which only the previous reconstructed frame is used to generate the side information. This operation results in a low-delay system since no complicated MCI performed at the decoder. The similar work can be found in [71] which sends the DCT hash and performs bidirectional hash motion estimation for side information generation. In [72], a low-delay system via employing an extrapolation technique for side information is proposed. In [73], authors discussed the solution using multiple side information from multiple reference frames. While in [74], authors explored the case of lossy side information. In [75], a modified three dimensional recursive search block matching is proposed to increase the accuracy of the side information. The simulation results showed the significant improvement over normal full search block matching algorithm in RD performance.

In [76], the mesh-based MCI is proposed to resolve the problems occurred in the side information interpolation based on block translational motion model. The proposed solution also addressed the problem of motion discontinuities and occlusions. The overall increasing accuracy of side information leads to the improvement in WZ coding performance. A new scheme with side matching during the frame interpolation is presented in [77] to improve the side information accuracy, in which the wrong compensated blocks are corrected by using side matching and bidirectional searching. In [78], the side information generation is considered in a rather complicated real environment, where the side information is extracted from the videos captured by a camera moving in a 3-D static environment. Especially, two interpolation methods based either on block matching along epipolar lines or on 3-D mesh fitting are proposed. To further cope with the problem caused by the assumption of linear motion between key frames, two kinds of feature point tracking techniques are also presented. In [79], an adaptive frame interpolation is proposed. The work proposed is based on a block-adaptive matching algorithm, in which the block size is adapted to the local activity within frame using block merging or splitting. The simulation proved that using adaptive frame interpolation performs better than using fixed size of the block during the interpolation. The work from [80] can be considered as an improved version

of the side information generation method of IST solution. It mainly added various block size during forward motion estimation and weighted distortion function during model decision and median filter to smooth the motion vector field. The PSNR improvement is reported around 0.4 dB over IST solution. In [81], side information generation using optimal filtering techniques is presented. The optimal filter is used to predict the MV between the side information and the WZ frame, which will be corrected by a conventional motion search in a decoded WZ frame. The side information interpolated by these refined MVs is fed into decoding process to provide a higher quality WZ frame.

2.5.6 Advances in Other Aspects

Some achievements in DVC research community regarding to other aspects in DVC are not covered by above sections. To help reader easily to understand the proposals of this thesis, some advanced research achievements are discussed in the review part of later chapters and introduced together with the proposed algorithms of this thesis.

Chapter 3 includes the review of techniques regarding to the side information refinement and rate control at the encoder or the decoder in WZ coding solutions. The recent advances in WZ coding applications e.g. multiview video coding will be addressed in Chapter 4. Another important application of WZ coding is robust video transmission, which will be studied in Chapter 5.

2.6 Further Discussions

With summary of research achievements presented in previous sections, it can be noted that most algorithms work on how to improve the compression performance of current WZ video coding from every possible directions. But few of them concern the influence of the decoder complexity caused by proposed research algorithms. Although in WZ video coding the decoder is supposed to be more complex, the complexity of the decoder cannot be infinitely large. The complexity of the decoder directly influences the delay of the system output, the efficiency usage of the resources etc. It is desired to optimise the decoder for current available WZ video coding solutions.

Some typical factors which directly influence the decoder complexity actually can be optimised to reduce the decoder complexity. Firstly, the size of the input block should be reduced since it causes serious delay of the decoding. Secondly, the current side information generation methods to provide the quality side information thus to achieve higher system performance is too complex. Moreover, the side information refinement process which causes several rounds of decoding is desired to be optimised. Finally, the feedback process involved in the request of

parity bits during the decoding should be designed in a more efficient way so that the repeat transmission, decoding and system output delay caused correspondingly can be minimized. The optimisation process should not degrade the system performance too much meanwhile.

Even though the IST TDWZ approaches to the conventional inter-frame coding e.g. H.263+ for certain video sequences, the performance of TDWZ generally cannot match that of the inter-frame coding based on the prediction structure. One of the major reasons is that the temporal correlation is never exploited in the above WZ video coding solutions. To further improve the performance of TDWZ and reduce its gap to that of the conventional inter-frame coding, removing the temporal redundancy among the WZ frames is a possible direction. However, suitable algorithm has to be chosen with keeping the encoder low complexity.

2.7 Chapter Summary

This chapter gives a general review of current WZ video coding from the basic theory background to several mature WZ video coding architectures proposed by different leading research groups. Particularly, the detail operations of each component in the architecture are introduced and individual experimental results are presented to make comparison. The advantages and disadvantages together with the limitations and possible applications of each solution are also discussed. Besides these leading research groups, other research advances proposed in the research community are also reviewed by grouping them into several categories including quantisation, transformation, channel coding and side information generation etc. Moreover, based on the review of current WZ video coding, possible research directions namely the motivations that this thesis is based on are discussed subsequently.

The next chapter will introduce the first contribution of this thesis which describes how to build an advanced pixel domain WZ video coding with optimised decoder.

Chapter 3

3. Pixel Domain Wyner-Ziv Video Coding with Optimised Decoder

In the current pixel domain video coding (PDWZ) solution, the decoder performs most complex tasks e.g. motion estimation (ME) and motion compensated interpolation (MCI). Most algorithms designing to improve system performance of PDWZ are operated at the decoder, which makes the decoder even more complex. However, the complexity of the decoder cannot be infinitely large because it influences the efficiency of resources' usage and system decoder output. The performance of current PDWZ coding is improved significantly by mutual efforts that research community made but fewer of them concern the complexity of decoder. In fact, the decoder of current PDWZ coding can be further optimised and the whole system still remains the same performance. This chapter presents four proposals, which are slice structure design, backward multiple reference frame interpolation, embedded refinement of side information and adaptive feedback optimisation, specifically targeting to optimise four components at the decoder which are input block of channel codec, side information generation, side information refinement and feedback process. The proposals are verified by simulations that it can make an efficient low-delay decoder but still remains same or even higher performance comparing to the current issues.

3.1 Slice Structure

3.1.1 Background and Motivation

In the PDWZ, there are mainly two ways to compose the input block to encode the WZ frame. In [3, 25] from STD research group, each pixel of WZ frame is scanned row-by-row and then is uniformly quantised. The quantised symbols formed an input block which is to be binarised and fed into turbo encoder for coding and transmission. In [33] from IST research group, the input block is composed by the bit-planes extracted from the quantised symbols. The frame is sent via the format of bit-plane. The performances of both structures are similar. It should come to attention that in both architectures, the size of input block is assumed arbitrarily large thus can deal with all bits from the whole frame or one bit-plane at once. However, the size of the input block has direct impact on the processing time of turbo encoding, interleaver design, and turbo decoding. Too large input block will produce significant computation latency during the encoding and decoding process. In other words, the system will not be able to provide a timely WZ decoded output due to the enormous Slepian-Wolf turbo coding and decoding delay. The input

block with arbitrarily large size is not feasible in practical applications. For example, a QCIF type WZ frame is quantised with 16 quantisation levels, thus the size of input block is $176 \times 144 = 25344$ bits in [33] and even larger as $176 \times 144 \times 4 = 101376$ bits in [25]. The input block could be times for other types of video sequences, i.e. CIF, SIF. In order to improve the decoding efficiency, the input block in PDWZ has to be reduced.

It is quite usual to divide the binary stream into several small blocks for coding and transmission, but in PDWZ coding, this possible solution is not feasible because of the “request-and-decode” process [3, 25] at the decoder. In PDWZ coding, the decoder estimates the overall symbol error probability of the whole WZ frame or a bit-plane transmitted to determine if further parity bits are needed for decoding. However, the overall symbol error will not be able to be estimated since the symbols cannot be reconstructed if only a small fraction of bits is received.

Moreover, the decoder applies the “request-and-decode” to perform rate control. When decoding is deemed not successful for the overall symbol error probability lower than certain threshold, the decoder will send the request back to the encoder for more parity bits. However, the encoder has no knowledge that where are the errors happened and which part of parity bits should be sent. It only can “blindly” send the parity bits according to the puncturing pattern determined by next lower coding rate. This causes the waste of transmission that some unnecessary parity bits are sent to help decoding some bits which are already correctly decoded. The decoding will still request more parity bits until the enough parity bits are received. It can be seen that by dividing the WZ frame into several small parts for transmission, the transmission of parity bits is not aiming to decode the whole frame but a small part of the frame therefore reduces the possibility to transmit the unnecessary parity bits.

In this section, a slice structure is proposed for PDWZ in order to cope with above two drawbacks. Firstly, the WZ frame is divided into slices for transmission. The input block is composed by all the binaried bits from quantised symbols of a slice, which has much smaller size than that composed of bits from a whole frame or a bit-plane. The encoding and decoding process thus can be much faster. Moreover, the WZ frame is sent slice by slice, the decoder will try to decode current slice first. If the decoding failed, the decoder will request more parity bits from the encoder until current slice is decoded successfully, then the next slice will be sent subsequently. One advantage to apply slice format is that even encoder “blindly” send the parity bits, the locations of decoding errors are limited into one slice other than a whole frame or a bit-plane. Thus the parity bits sent are more accurately. For example, in normal PDWZ, the decoding is probably failed only due to a small high motion area in the frame. The decoder will stop requesting parity bits only when the parity bits specially used to protect this high motion area

arrived. Before that, all the parity bits sent are wasted. However, in the proposed slice architecture, the high motion area caused failure of decoding probably only belongs to one slice. The decoder requests more parity bits only when this slice is sent. Before that, for other slices like slices only contain the background information, the decoding will be deemed successful thus no requests for more parity bits will happen. With slice structure, both drawbacks mentioned above can be solved and the system performance can be further improved. The following part explains the whole architecture of PDWZ with slice, in which the mechanism of how to decide the slice size is also shown.

3.1.2 Architecture and Operations

Fig.3.1 shows the overall architecture of the proposed PDWZ with slice structure(PDWZS), whose backbone is mainly inherited from PDWZ in [25, 33]. In the proposed PDWZS, the WZ frame sequence is composed by each even frame X_{2i} of video sequence, and each two adjacent odd frames X_{2i-1} and X_{2i+1} are used as key frames, where i is the frame temporal index of sequence. The WZ frame is intra-frame coded at the transmitter and inter-frame decoded at the receiver with the help of side information, denoted by Y_{2i} .

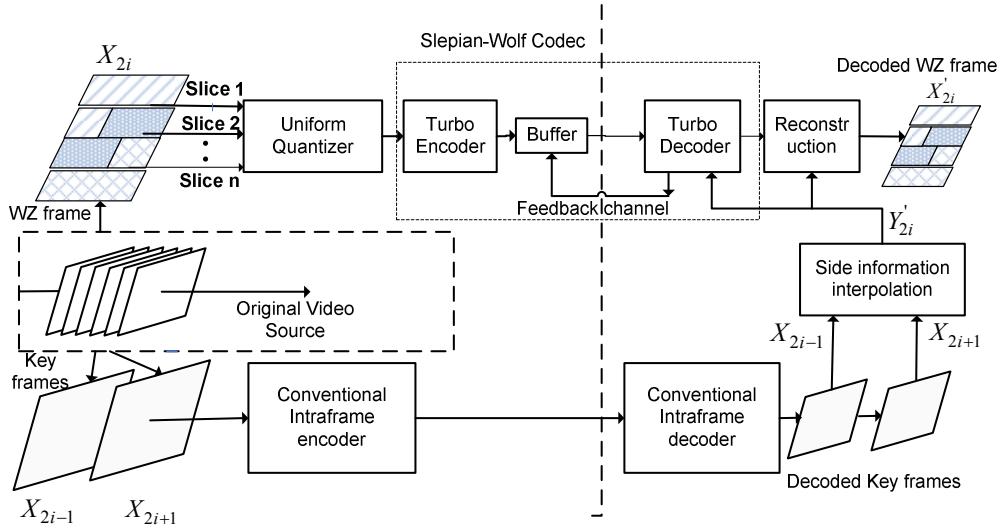


Fig.3. 1 The architecture of proposed PDWZ with Slice structure

Same as in [25, 33], the key frames are transmitted via conventional intra frame video coding, which are assumed to be transmitted and reconstructed perfectly at the receiver in the PDWZS. The side information Y_{2i} is interpolated from the received key frames by using motion compensated frame interpolation [33].

The WZ frame is divided into slices for transmission. Each pixel in the slice is quantised by using a uniform scalar quantiser with $2^M = \{2, 4, 8, 16\}$ levels. The quantised symbols then are binaried and fed into Slepian-Wolf codec for coding. Here, 3GPP standardised turbo codec with

polynomial generator of (13,11) and constraint length of 4 is implemented as the Slepian-Wolf codec. Together with the rate compatible puncturing method [27] it can achieve flexible code rate to adapt the statistical change of side information Y_{2i} . The turbo coded parity bits are saved into a buffer for transmission and all the systematic bits are discarded.

At the receiver, Laplacian distribution is used to model the statistic dependency between the original WZ frame X_{2i} and side information Y_{2i} . The Laplacian parameter is derived by using side information Y_{2i} and one of the adjacent key frames, with which the slice will be decoded by turbo decoder with combination of the received parity bits, side information Y_{2i} and Laplacian distribution parameters. During the decoding of each slice, the “request-and-decode” process will initially request a fraction of parity bits from the buffer of the encoder to commence the decoding. A slice is considered successfully decoded when the resulting bit-error-rate (BER) is lower than a predefined threshold, e.g. 10^{-3} and the next slice is sent for decoding subsequently. Otherwise, a request for more parity bits is sent back to the encoder via feedback channel. This process is repeated until the BER of the current slice decoding is lower than 10^{-3} and then next slice is sent subsequently. Reconstruction of a complete WZ frame is performed after all successive slices received.

This part illustrates the criteria of defining the slice. There can be many issues regarding to how to divide the frame into slices. For example, the slice can be obtained by uniformly dividing the frame into slices with same length and size. Or as in H.264 or MPEG7, the slice can be composed according to the content of frame, namely the similar content in the frame will be grouped into same slice. However, this kind of division has two disadvantages: a content detection method has to be devised, which will increase the computation complexity at the encoder. Moreover, some slices such as slice containing most background could be too large for the turbo encoder. In many practical turbo codec, the interleaver size is limited. Hence, we could design the slice according to the size of interleaver in turbo codec. In the PDWZS, we purposely apply a practical 3GPP turbo codec, which has limitation of the interleaver size between 40~5114 bits. With this limitation, the size of input block is also limited to 40~5114 bits, which will be used as criteria to divide the slice in the PDWZS. The process of how to define the slices is described as following.

With 3GPP turbo codec, the size of a slice after quantisation cannot exceed 5114 bits. Given the quantisation level is 2^M and MB size is b , then the maximum length l of each slice can be obtained by:

$$l = \left\lfloor \frac{5114}{b^2 * M} \right\rfloor \quad (3.1)$$

where $\lfloor x \rfloor$ denotes the biggest integer smaller than or equal to x . Therefore, a WZ frame with r rows and c columns can be divided into n slices with the length of l , plus a remaining slice with the length of l_1 where $0 < l_1 < l$. The relationship between n and l_1 is expressed as:

$$(n, l_1) = \text{Quo_rem}\left(\frac{r * c}{b^2 * l}\right) \quad (3.2)$$

The function “*Quo_rem*” returns quotient of the division to n and the remainder to l_1 . For an example, with 8-levels quantisation and MB size of 16x16, a QCIF frame will be divided into 16 slices with length of 6 MBs and 1 slice with length of 3 MBs. Therefore, the total number of slices to be coded and transmitted is $n+1$ (or n where $l_1 = 0$). Note that the following case ($l_1=1$, $b=4$ and $2^M=2$) is invalid since the minimum size of the input block is 40 bits.

3.1.3 Results and Discussions

In the simulation, the proposed PDWZS and a normal PDWZ(similar framework as in [33] and indicated by “PDWZ”) without slice structure are simulated. H.263+ inter-frame coding with IBIB structure and H.263+ DCT intra-frame coding are listed as references. The first 101 frames of QCIF video sequences “Foreman” and “Mother and daughter” are investigated in the simulation. The frame rate for both sequences is 30fps thus even frame (WZ) rate is 15 fps. Each pixel is quantised with levels of $2^M = \{2, 4, 8, 16\}$ to give four rate-distortion points. The side information is interpolated by normal bidirectional motion compensated interpolation from two adjacent frames without any further refinement process e.g. spatial smoothness. The micro block (MB) size is set to 16x16 and full search ME window size is ± 7 pixels and the bidirectional ME window size is ± 4 pixels.

Fig.3.2 and Fig.3.3 show the RD performance for above solutions. It can be observed that all PDWZ solutions outperform the H.263+ intra-frame coding but still have large gap to the H.263+ inter-frame coding. Comparing with the normal PDWZ, it is observed that system performance is further enhanced by PDWZS. The PSNR gain can outperform up to 1.6 dB for “Foreman” sequence and 1.4 dB for “Mother and daughter” sequence. The system improvement achieved by PDWZS largely attributes to the optimisation of the “request-and-decode” process by slice structure. WZ frame is decoded slice-by-slice and only the slice which is not successfully decoded requires more parity bits to be sent. This operation avoids sending unnecessary parity bits thus the rate is tightly controlled.

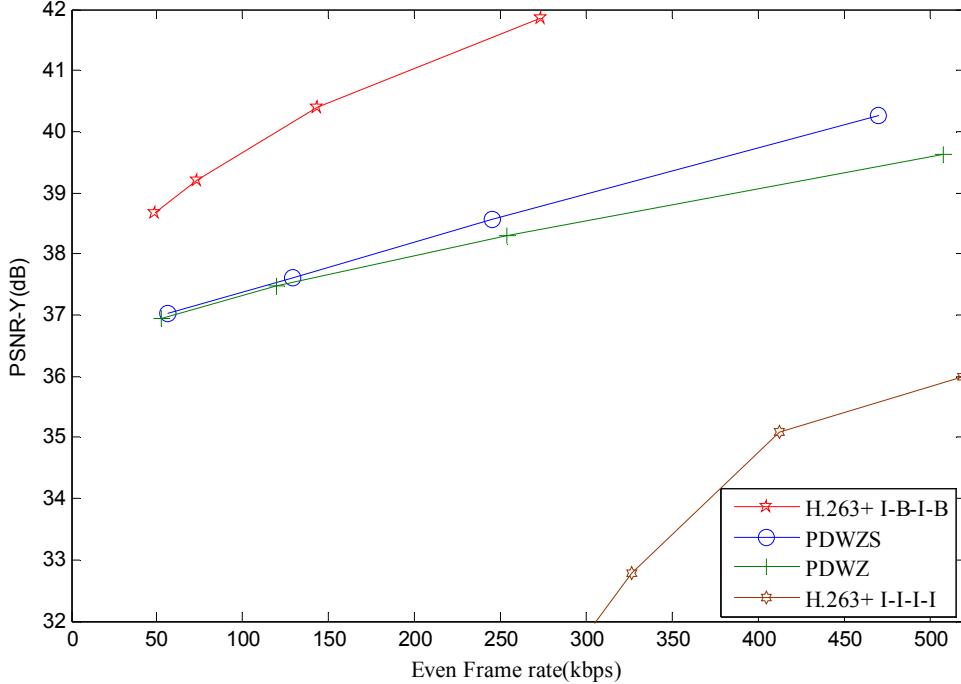


Fig.3.2 PSNR performance of PDWZS for Foreman sequence

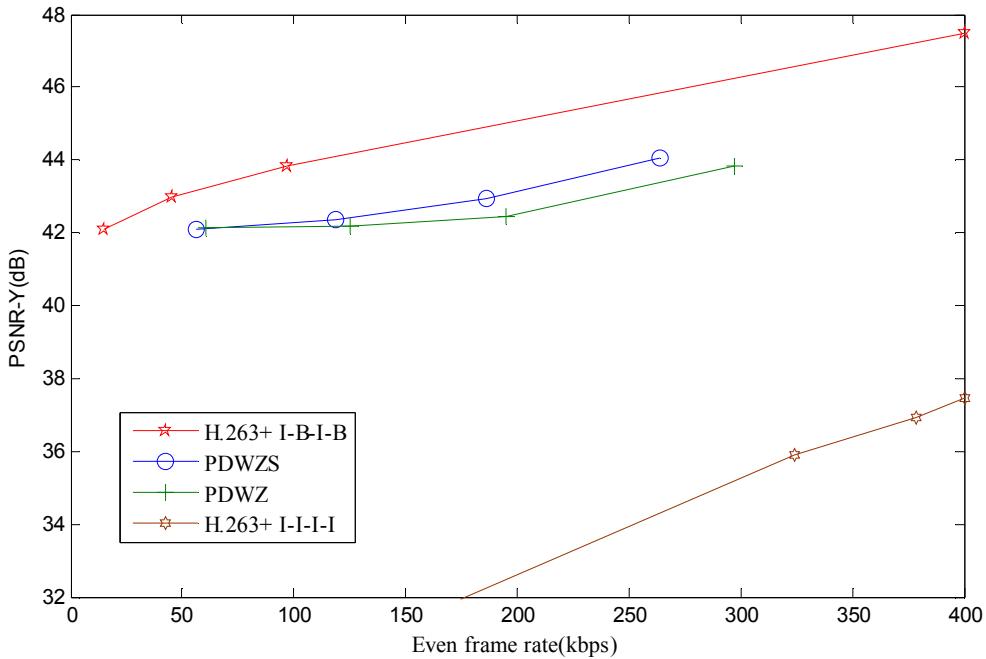


Fig.3.3 PSNR performance of PDWZS for Mother and daughter sequence

Fig.3.4 shows the perceptual quality comparison for Foreman sequence between PDWZS and normal PDWZ without using slice. Given the same bit rate of WZ frame, the amount of parity bits sent is thus fixed. It is apparent in the picture that the PDWZS efficiently utilises received parity bits and corrects most error. While for normal PDWZ, many received parity bits are wasted so many errors still can be spotted and need further more parity bits to correct the errors.



(a) PDWZ without slice
(b) PDWZ with slice
Foreman 53rd frame 16 Quantisation level 350kbps

Fig.3. 4 Perceptual performance of PDWZS and normal PDWZ

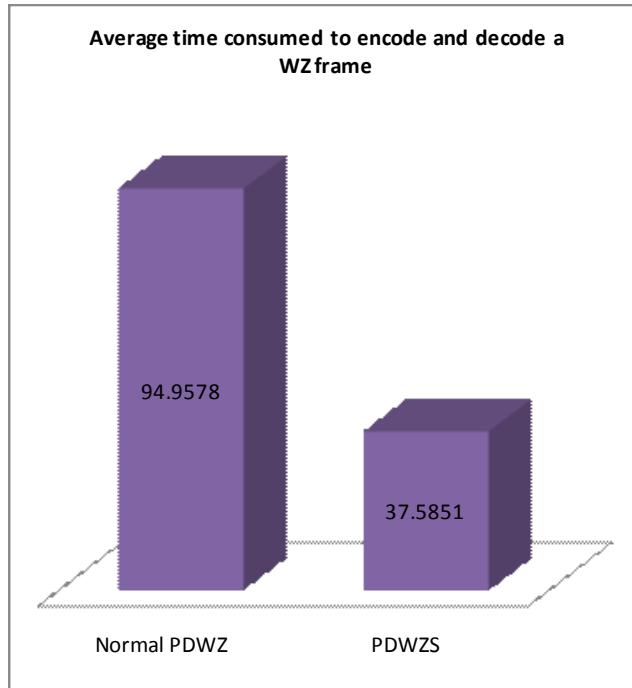


Fig.3. 5 Average time (seconds) consumed comparison between PDWZS and normal PDWZ

More importantly, the PDWZS is processed much faster than normal PDWZ due to the design of small input block. In the simulation, 50 WZ frames are tested between PDWZS and the normal PDWZ without using slice. During the simulation, the processing time is specially recorded. The simulation is running in the environmental of Matlab 7.0, windows XP on the computer with Pentium4 2.66GHZ, 2GRAM.

Fig.3.5 shows the average time (seconds) consumed to encode and decode a WZ frame with 8 quantisation levels. It is evident that the PDWZS saved more than 60% of processing time thus is more efficient than the normal PDWZ.

3.2 Backward Multiple Reference Frame Interpolation

In WZ coding, the accuracy of side information directly influences the rate distortion performance of system. Here the word “accuracy” reflects the similar degree between the original WZ frame and side information interpolated. The more accurate is side information, the more accurate is the Laplacian prediction in decoding. Therefore less parity bits are needed to decode the WZ frame. In Chapter 2, several kinds of frame interpolation techniques were introduced generally. Among these interpolation techniques, the bidirectional MCI introduced in [33] is superior to the rest algorithms to produce the side information with high quality. It is evident that the bidirectional MCI leads to further enhancement in R-D performance of WZ video coding in [31, 33].

However, the computation is also very high for bidirectional MCI due to the bidirectional motion search inside. This is another factor that increases the decoder complexity. In PDWZ of IST or STD, further operations such as motion smoothness etc. will also be applied additionally after bidirectional MCI. Too much time consumed on side information generation will cause huge delay of WZ frame decoding output. This section mainly focuses on how to optimise side information generation process. A new MCI algorithm is proposed based on a fast and low computational Backward Multiple Reference Frame ME. We refer this MCI method as BMRF and bidirectional MCI as BID through the whole thesis. The target of BMRF is to offer relatively better side information than the average interpolation or simple temporal linear MCI (TLIN), but not to suffer huge computation as BID. The quality of side information generated by BMRF usually is inferior to that generated by BID, thus the system performance is degraded comparing to that of the system using side information generated by BID. However, an embedded refinement technique will be introduced in next section to compensate the quality lost of side information for BMRF. It is proved in Section 3.3 that using BMRF plus an embedded side information refinement can be more effective than purely using BID. In this section, we mainly focus on introducing the algorithm of BMRF and concerning the quality of side information. The performance of WZ video coding with BMRF will be revealed in Section 3.3. Then a review of most frame interpolation techniques will be presented first and then the detail of BMRF algorithm will be described. Later section presents the simulation results and discussions.

3.2.1 Review of Frame Interpolation

The earlier frame interpolation algorithms do not consider moving objects during frame interpolation. Two representative algorithms are frame repetition and temporal frame averaging in this category. In term of performance, these two algorithms cause jerkiness and blurring effects

on moving objects in interpolated frame, thus degrade the interpolated quality rapidly.

Later frame interpolation algorithms consider moving objects during frame interpolation. These algorithms are known MCI employing ME between two consecutive frames. Since MCI detects the moving objects, therefore the interpolation quality can be improved significantly. TLIN is a basic and conventional MCI method based on block ME (BME). It is fast and easily implemented but could easily produce hole and overlapped regions in the interpolated image during the interpolation since the estimated motion vector (MV) will not always intercepts the centre of the interpolated macroblock (MB).

In order to deal with above block artifacts inevitably introduced by the BME, further improvements have been explored in this direction. One way is to refine the MVs of the MB in interpolated image. In [82], a new MCI method based on bidirectional ME, namely BID, has been proposed. This algorithm employs exhaustive bidirectional ME to refine the bidirectional motion vector and eliminates the holes or overlapped regions in the interpolated frame. The interpolation quality is much better than that of TLIN, but the computation is very high due to the exhaustive bidirectional motion search. In [83], the weighted method has been used to improve the accuracy of the estimated MVs of the current MB by combining the MVs of neighboring sub-MBs. There are some other algorithms like weighted median filter and spatial-temporal smoothness making correction on MVs errors and the results are also positive in [83-85]. Another way to improve interpolation quality is to employ more effective ME algorithms. In [86], Overlapped Block ME (OBME) is utilised to achieve the higher accuracy of MV, while in [87, 88], the multiple motion and non linearity motion have been considered. Moreover, the sender assisted methods have been used by sending the information like MVs or DCT coefficients of the skipped frame to aid interpolation in [70, 89]. However, this kind of method increases transmission bandwidth and puts much computation at the sender. It can be seen that all these algorithms mentioned above only use two consecutive frames to perform interpolation, no algorithms so far in this research area have considered using multiple frames to performance interpolation.

Unlike most MCI algorithms, BMRF builds the connection among multiple frames using backward multiple reference frame ME, which is introduced in [90-92], to perform weighted frame interpolation. It is found in this thesis that the BMRF is extremely effective in reducing the MCI computation complexity without sacrificing much of the interpolation quality. The BMRF algorithm is explained in detail in the following section.

3.2.2 BMRF

In order to help understanding the BMRF algorithm, we denote an adjacent video frame

sequences as $f(x,1), f(x,2), \dots, f(x,k)$, where x is the MB coordinate and k is the frame index. Given that $1 < n < k$, we assume the frame $f(x,n-0.5)$ is to be interpolated by 3 adjacent received key frames denoted as $f(x,n-1), f(x,n)$ and $f(x,n+1)$ namely previous, current and future frame respectively. Assuming that v_i denotes motion vector (MV) from frame $f(x,n+1)$ to frame $f(x,n)$ and v_j denotes MV from frame $f(x,n)$ to $f(x,n+1)$ where $f(x,n)$ and $f(x,n+1)$ are the reference frames for above two MEs, respectively. $f_R(x,n)$ means intensity value of a MB positioned at x in frame $f(x,n)$. The next part will discuss the process of BMRF performing weighted frame interpolation on $f(x,n-0.5)$ frame as illustrated in Fig.3.6

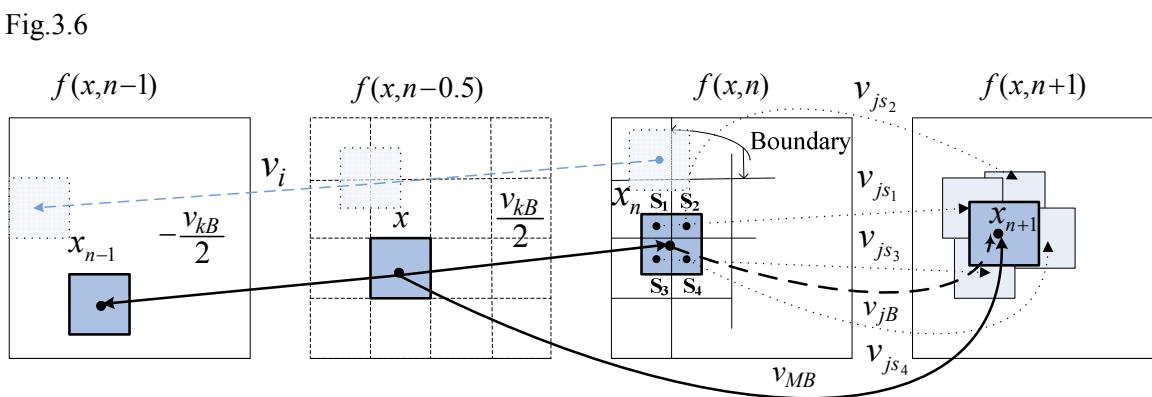


Fig.3.6 Weighted Frame interpolation in BMRF

Firstly, the backward BME needs to be performed to obtain v_i and v_j . The reason why it is called backward BME is that the future received frames have been used as a reference frame, which is inverse to the traditional video coding where normally the previously received frames are used as reference frames in motion compensation. Frame $f(x,n-0.5)$ lies on the half way between two frames in term of time. If frame interpolation is carried out at this point, this is known as TLIN as expressed as:

$$f(x + 0.5 * v_i, n - 0.5) = 0.5 * (f(x, n - 1) + f(x + v_i, n + 1)) \quad (3.3)$$

Unfortunately, the hole and overlapped regions will appear in the interpolated image since not all MVs intercept the centre of each MB of $f(x,n-0.5)$. In order to eliminate these block artifacts, a bidirectional MV assuming intercepting centre of each MB of $f(x,n-0.5)$ needs to be initialised. After the first BME to obtain v_i , each pixel in the interpolated frame, apart from the pixel located in the hole region, will have a motion vector determined by the $MV \in v_i$ of the MB. The initial bidirectional MV of each MB of $f(x,n-0.5)$ will be determined by MV of the centre pixel of the current MB. If the centred pixel does not have any MV due to the hole region, the nearest pixel to the centred pixel with MV will be chosen. Assume now $f(x,n-0.5)$ has bidirectional MV v_k , then $f(x,n-0.5)$ can be interpolated bidirectionally as

$$f(x, n - 0.5) = 0.5 * (f(x - 0.5 * v_k, n - 1) + f(x + 0.5 * v_k, n + 1)) \quad (3.4)$$

Although the block artifacts are eliminated by above process, the quality of interpolated image is still not acceptable due to the poor accuracy of initial bidirectional MV assigned. Based on this, many algorithms such as BID [82] and WAMCI [83] etc. have been used to refine this initial bidirectional MV. But an apparent fact is that massive computation will be raised. BMRF will not refine the initial bidirectional MV but extend the trajectory of this MV to further received frame by using multiple reference MEs, and to find another matching MB in other frames to perform weighted frame interpolation with additional frame as illustrated in Fig.3.6. Assume the MB positioned at x in frame $f(x, n-0.5)$ has the initial bidirectional MV v_{kB} , two corresponding MBs at each end of the MV trajectory between two adjacent frames can be located as:

$$x_{n-1} = x - 0.5 * v_{kB}, \quad x_n = x + 0.5 * v_{kB} \quad (3.5)$$

Then a weighted method is applied to compose v_{jB} based on the calculation of v_j to find the corresponding x_{n+1} in $f(x, n+1)$ for the current x_n in $f(x, n)$.

As shown in Fig.3.6, generally, x_n is not aligned with the MB boundary but divided into four parts: s_1, s_2, s_3 and s_4 by the MB boundary. Assuming the previously calculated MVs from $f(x, n)$ to $f(x, n+1)$ of MBs located in s_1, s_2, s_3 and s_4 are $v_{js1}, v_{js2}, v_{js3}$ and v_{js4} respectively, the motion vector v_{jB} can be composed by following weighted method:

$$v_{jB} = \frac{v_{js1}S_1 + v_{js2}S_2 + v_{js3}S_3 + v_{js4}S_4}{S_1 + S_2 + S_3 + S_4} \quad (3.6)$$

According to v_{jB} , the position of the corresponding x_{n+1} can be obtained as:

$$x_{n+1} = x_n + v_{jB} = x + 0.5 * v_{iB} + v_{jB} = x + v_{MB} \quad (3.7)$$

Then, the current MB can be interpolated by,

$$\begin{aligned} f_B(x, n - 0.5) &= \frac{f_B(x_{n-1}, n - 1) + f_B(x_n, n)}{4} + \frac{f_B(x_{n-1}, n - 1) + f_B(x_{n+1}, n + 1)}{4} \\ &= \frac{f_B(x_{n-1}, n - 1)}{2} + \frac{f_B(x_n, n) + f_B(x_{n+1}, n + 1)}{4} \end{aligned} \quad (3.8)$$

The above process will be performed MB by MB until a whole frame is interpolated. Then v_j will be save to cache to be used as the first ME to interpolate the next frame $f(x, n+0.5)$. Note that BMRF does not carry out exhaustive search and refinement of MV whilst interpolating a frame. Instead, BMRF uses frames $f(x, n-1), f(x, n)$ and $f(x, n+1)$ to perform weighted interpolation of the frame $f(x, n-0.5)$. In this way BMRF reduced the computation complexity tremendously as compared to BID and WAMCI.

3.2.3 Results and Discussions

The proposed BMRF is tested with common QCIF sequences such as “carpone”, “foreman”, “costguard” and “salesman”, each of them represents a different kind of video. A total of 101 adjacent odd frames have been picked out to perform the interpolation. The simulation uses three MCI algorithms (TLIN, BID and BMRF) to interpolate 100 frames which are compared to the original frames in order to calculate the PSNRs. An average PSNR is calculated to show the performance of BMRF in comparison to TLIN and BID. Furthermore, the original 86th frame and the interpolated versions of this frame rebuilt by those three MCI algorithms are presented to show the perceptual quality. Finally, the computational complexity of each algorithm is analyzed.

For the frame interpolation process, video frame is divided into block size of 16x16, hence there are 99 MBs per frame. The backward BME search step is set to 7 bits for all MCI algorithms and the bidirectional ME refinement search step is set to 4 bits which is only used by BID.

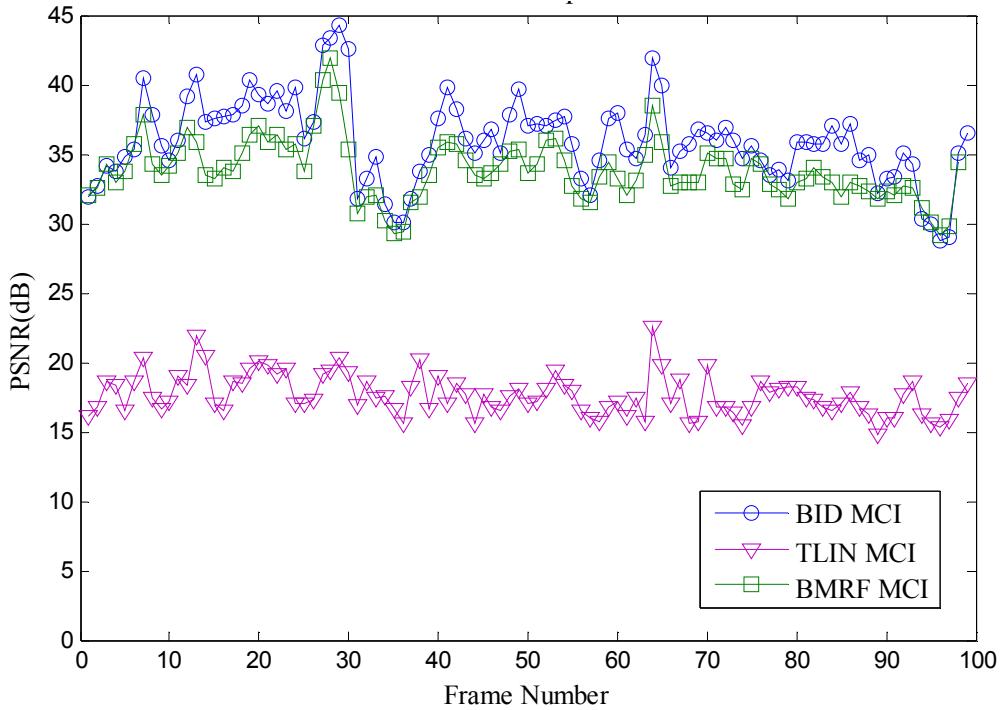


Fig.3. 7 PSNR comparison for TLIN, BID and BMRF interpolation for Foreman sequence

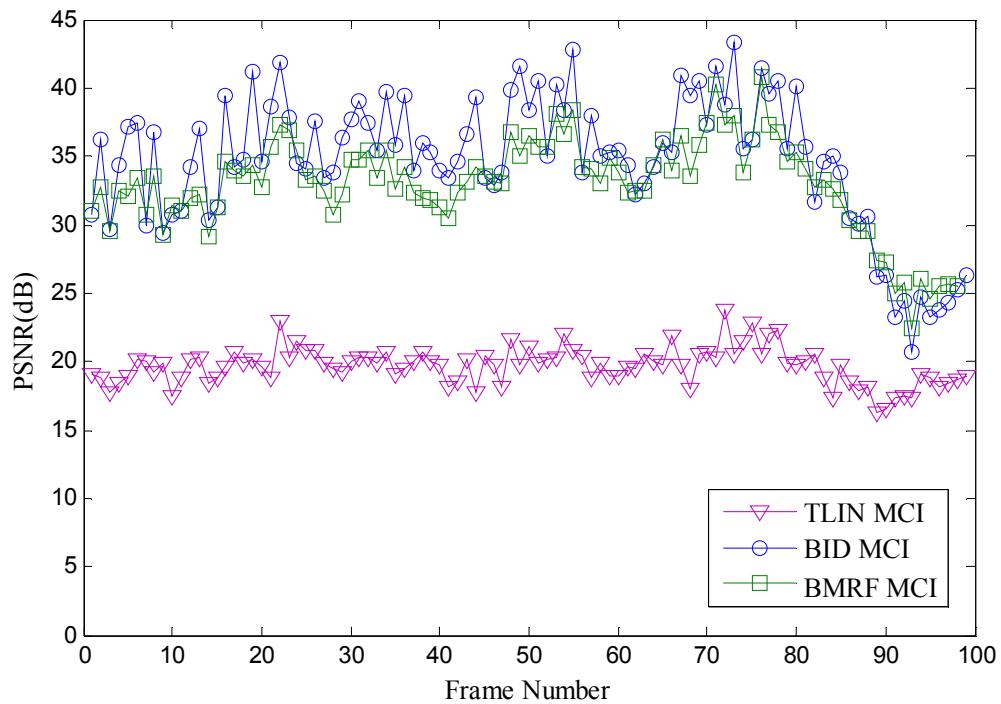


Fig.3.8 PSNR comparision for TLIN,BID and BMRF interpolation for Carphone sequence

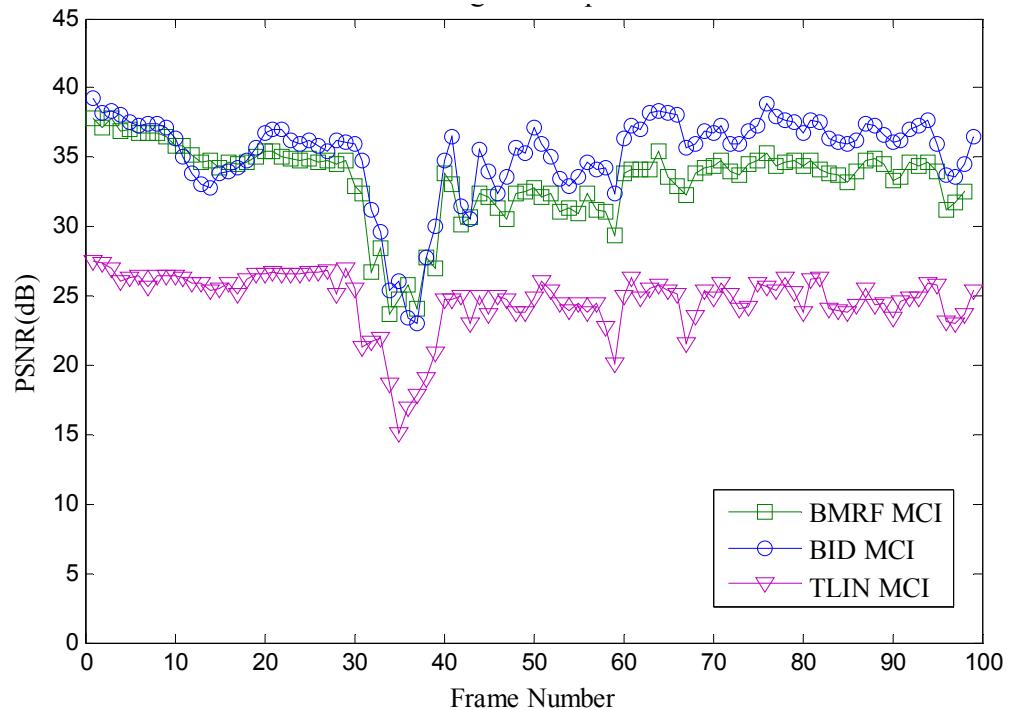


Fig.3.9 PSNR comparison for TLIN,BID and BMRF interpolation for Coastguard sequence

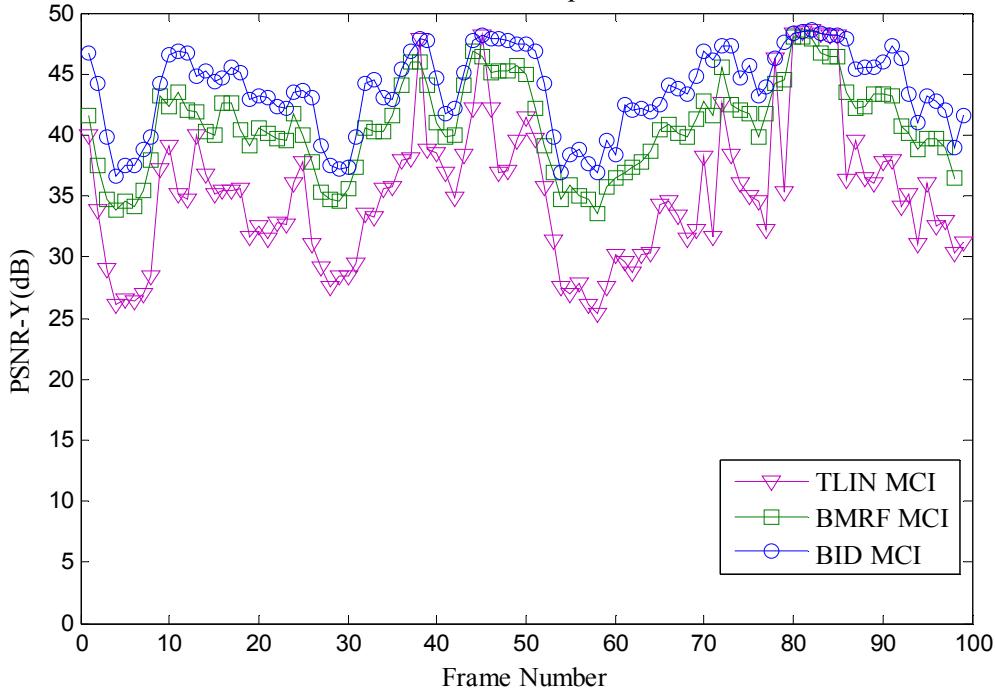


Fig.3. 10 PSNR comparison for TLIN,BID and BMRF interpolation for Salesman sequence

As a result of the simulation, Fig.3.7, Fig.3.8, Fig.3.9, Fig.3.10 and Table 3.1 showed that, in most circumstance, the BID provides superior interpolated quality and PSNR gain followed by BMRF and TLIN. On the other hand, BMRF has average of 1.64 dB PSNR gap to the BID. Nevertheless, the rebuilt interpolated image showed very similar perceptual quality between the BMRF and BID as shown in Fig.3.11. Obviously, BMRF and BID are much better than the simple TLIN in both PSNR and perceptual quality. Notice that although the TLIN algorithm can track the moving objects, it produces overlapped and hole regions.

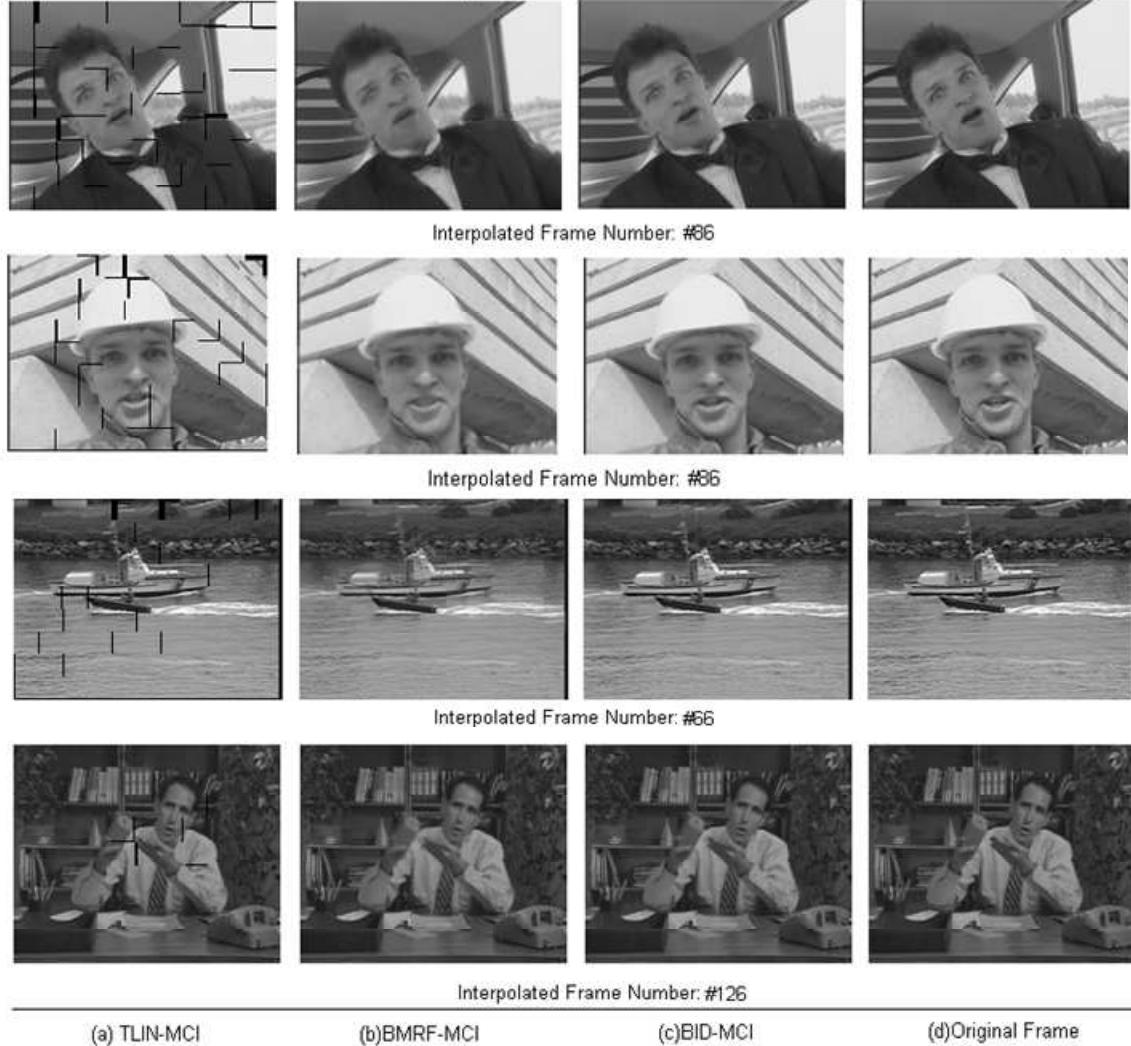
Table 3. 1 PSNR performance of different MCI algorithms

PSNR-Y(dB)	TLIN	BMRF	BID
Carphone	19.47	32.68	34.36
Foreman	26.28	34.03	35.96
Coastguard	24.64	33.31	35.22
Salesman	35.18	42.69	43.74

Table 3.2 shows the number of search MBs required to interpolate one frame for TLIN, BMRF and BID. Apart from the ME between two adjacent frames, TLIN only needs 2 reference MBs to interpolate one block. The proposed BMRF only needs 3 reference MBs; meanwhile, the additional MVs (v_j) can be reused as the first ME for the next frame interpolation thus computations is saved. The BID computation varies depending on the bidirectional ME

Table 3.2 Computational complexity of different MCI algorithms

	TLIN	BMRF	BID(d=4)
No. of search MBs required for interpolation of one frame	198	297	7920
Average time(second) for interpolating a frame	0.022(s)	0.036(s)	0.978(s)

**Fig.3.11 Perceptual quality comparison for TLIN, BID and BMRF interpolation**

refinement search step, which gives the number of searching MB for each interpolated MB by $(2d + 1)^2 - 1$. It can be seen in the table that the computational complexity of BMRF is just a slightly higher than that of TLIN but is much lower than that of BID. For a total number of block to be interpolated is 99, TLIN needs to search $99*2=198$ MBs and BMRF needs $99*3=297$ For

BID with $d=4$, the computation of the proposed BMRF is only around 3.75% of BID. This argument is further proved by the observed average time of interpolating a frame for above 4 sequences based on a standard PC (P4 2.66GHZ, 1G RAM).

The new BMRF MCI algorithm proposed in this thesis uses multiple frames ME to perform weighted frame interpolation. The simulation results show that BMRF can produce similar interpolation quality but largely reduce the computational complexity up to 92% comparing to that of BID MCI. Considering its interpolation quality and computation advantages, the BMRF is a sub optimal choice for side information interpolation in DVC. In next section, the performance of DVC with side information interpolated by BMRF is revealed. However, as expected since the interpolation quality of side information is inferior to that interpolated by BID, the performance of the DVC with BMRF will not show any advantage. However, in next section, another embedded refinement inside the turbo codec is proposed to compensate the quality loss of the side information generated by BMRF. It is proved that using BMRF plus an embedded refinement technique can be more effective than purely using BID for DVC, thus provide more superior performance.

3.3 Embedded Side Information Refinement

It has been emphasised lot in previous sections that how important the side information is in DVC. With the further research into WZ video coding, many researchers have realised that whichever interpolation method applied, the real content of the encoded WZ frame remains unknown during the interpolation. This implies all the current interpolation methods are not reliable since they are based on an assumption of smooth motion or linear translation inside the frame sequences. The interpolation quality could be degraded more when interpolating video with fast or irregular motion objects. However, in order to diminish the effect from above disadvantage and get further improvement of side information, several refinement methods applied to the decoded WZ frame have been explored in [68, 69, 73, 93]. The general principle is that a WZ frame is first decoded by using initial side information, and then it is refined by performing further MCI between two adjacent key frames again and used as the improved version of side information to perform the next round of n -iteration turbo decoding process. Thus final output is improved correspondingly. In next following sections, some existing refinement techniques are reviewed first and the embedded refinement technique of side information is proposed. Later sections bring the simulation results and discussions.

3.3.1 Review of Side Information Refinement

In [69], authors proposed the method called Motion Compensated Restoration(MCR) to aid the

refinement process of the side information. The block in the side information to be refined will search the best match block in temporal adjacent key frames, average interpolation frame and decoded WZ frame. It is proved in [69], the refinement process has achieved around 0.15dB gain in PSNR performance. In [36, 73], the refinement of side information is performed based on bit-plane transmission. The frame can be approximately reconstructed with received several bit-planes, then refined by performing MCI between key frames.

Subsequently, this refined WZ decoded frame is used as improved side information and feedback to turbo decoder again to help decode next bit-plane. The major difference between [33] and [36] is that [69] presented the refinement in block level. Only blocks in decoded frame matching the designed criteria are eligible to be refined, while [73] added chrominance information to aid the refinement of side information.

It is reasonable to refine the decoded WZ frame and use it as improved side information to achieve further system enhancement. However, it should be noticed that all above algorithms perform the refinement of side information outside the turbo decoder. After each round of refinement, turbo decoding needs to be operated again with whole round of n -iterations. Turbo decoding is a time consuming process especially for relative larger size of bit stream. This kind of refinement definitely times the decoding time and increased the complexity of the system. This section will bring a new refinement process for PDWZS, in which the refinement process is shifted into turbo decoder and performed synchronously after each iteration of turbo decoding. With this embedded process, the turbo decoding only needs to be operated once thus computation is greatly saved comparing to above refinement techniques. Moreover, in PDWZS, the refinement is processed to the slice of the side information. As mentioned above, it is not necessary to use a very complicated side information interpolation technique since no interpolation technique is reliable, the system thus need not spend much computation on the initial side information interpolation but use BMRF introduced in Section 3.2 to provide side information with relatively good quality. The performance of the whole system will be improved effectively by the embedded refinement process, and thus the quality lost will be compensated during initial side information interpolation.

3.3.2 Architecture and Operations

In Fig. 3.12, the architecture of PDWZS with embedded refinement is shown. Same as in Section 3.1, WZ frame sequence is composed by each even frame X_{2i} of video sequence, and each two adjacent odd frames X_{2i-1} and X_{2i+1} are used as key frames; the WZ frame is intra-frame coded at the encoder and inter-frame decoded at the decoder with the help of side information.

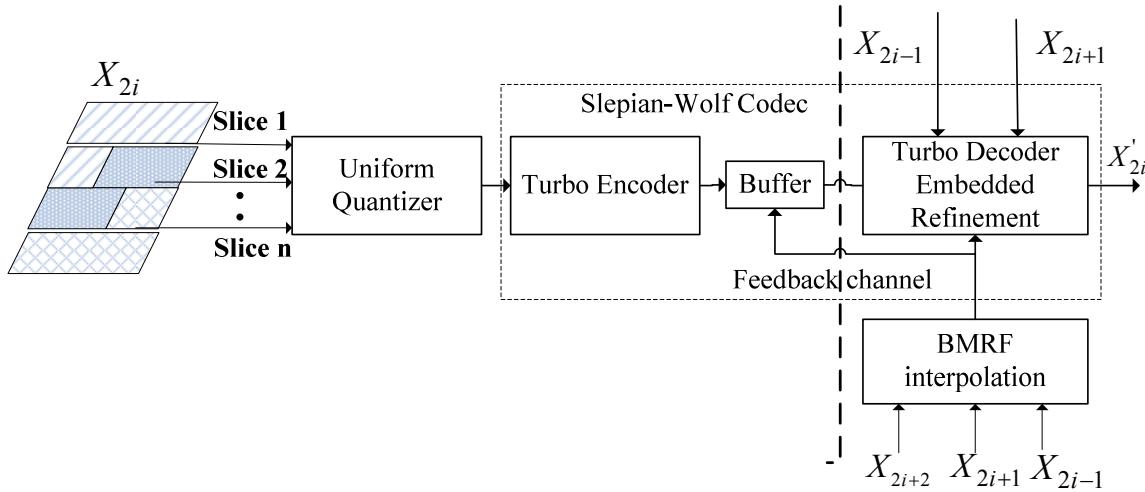


Fig.3. 12 The architecture of PDWZS with embedded refinement

In order to simplify the problem, the key frames are assumed transmitted and reconstructed perfectly at the decoder. The system starts firstly dividing current WZ frame into slices as specified in Section 3.1. Each slice will be uniformly quantised and sent to 3GPP turbo encoder for encoding. Only if the current slice is successfully decoded at the decoder, the next slice will be sent subsequently. At the decoder, the side information Y_{2i} is initially interpolated by using BMRF with the received key frames, and then it will help turbo decoder to perform the 1st iteration decoding. After each iteration of turbo decoding, the slice is reconstructed and then the embedded refinement technique refines the reconstructed slice block by block by Weighted Motion Compensation Refinement (WCMR). Afterwards, the refined slice is used as improved side information to help the next iteration of decoding. When all iterations are finished, if the turbo decoder still cannot reliably correct the major errors, more parity bits are requested from encoder via feedback channel until current slice is decoded with acceptable errors. The final WZ frame is reconstructed until all slices are received.

Fig.3.13 shows the details of the embedded refinement inside the turbo decoder. After the 1st iteration of decoding, some errors can be corrected by turbo decoder. Then, the output likelihood ratio of second SISO decoder is deinterleaved and hard quantised to compose decoded bits stream. An approximate decode slice is reconstructed. In general, this decoded slice is more accurate than initial slice of side information since some interpolation errors are corrected. Next, the WMCR is performed as described below.

Let $S_{2i}'(x,y)_B$ denote the MB located at (x,y) in S_{2i}' , X_{2i-1} and X_{2i+1} denote two adjacent key frames. Suppose symmetric motion vectors used for initial interpolation (that is BMRF) are v_1 (to key frame X_{2i-1}) and v_2 (to key frame X_{2i+1}). Based on v_1 and v_2 , the full block motion search in X_{2i}

X_{2i} and X_{2i+1} within a relatively small window size are performed. Note that ME search at this time

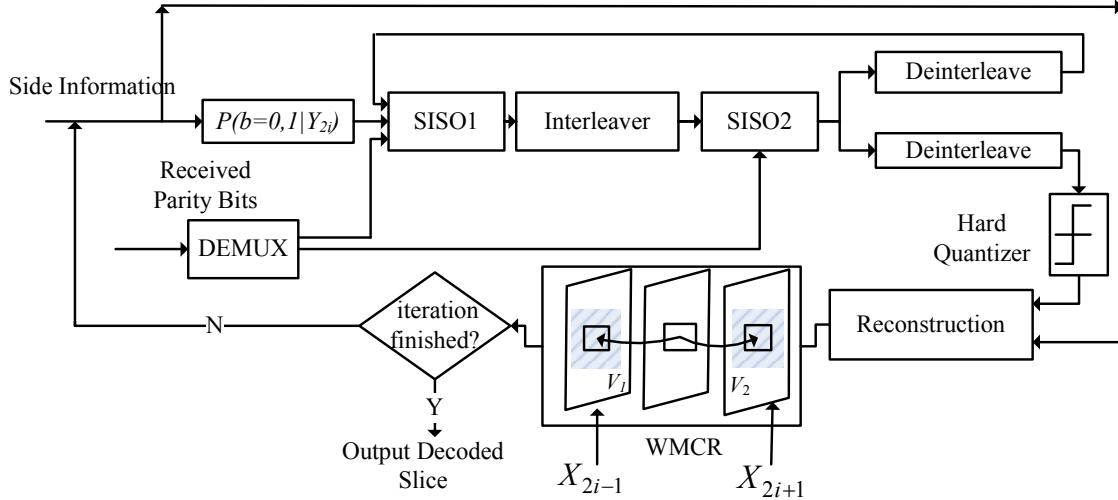


Fig.3.13 Embedded refinement inside the turbo decoder

is not necessarily symmetric any more like as done in bidirectional motion search. Since the decoded WZ frame is more accurate than side information due to the recovery of several errors, ME can be more effective and efficient than that of the bidirectional search in BID. After ME process, two blocks K_{1B} and K_{2B} with smallest SAD_1 and SAD_2 where SAD denotes Sum of Absolute Difference in X_{2i-1} and X_{2i+1} are located respectively. Then current block is interpolated and refined by the WMCR. We first introduce the concept of block accuracy from [83], which is defined as:

$$\sigma = \frac{1}{SAD} \quad (3.9)$$

Smallest SAD implies higher accuracy of block search. Then the current block is refined as:

$$\begin{aligned} S_{2i}'(x, y)_B &= \frac{\sigma_1}{\sigma_1 + \sigma_2} K_{1B} + \frac{\sigma_2}{\sigma_1 + \sigma_2} K_{2B} \\ &= \frac{SAD_2 * K_{1B} + SAD_1 * K_{2B}}{SAD_1 + SAD_2} \end{aligned} \quad (3.10)$$

The refinement process continued block by block until the slice S_{2i}' is refined. Then this refined S_{2i}' is used as an improved version of side information and fed into decoder again to help the next iteration of decoding. Note that Laplacian parameter a should be refreshed corresponding to the statistics change in side information. This operation is repeated inside the decoder until the whole decoding is finished.

3.3.3 Results and Discussions

In order to evaluate the performance of the proposed system, we simulated four types of PDWZ video coding systems:

- i) Normal PDWZ with BID side information interpolation and without refinement process:
The normal PDWZ has same architecture as IST PDWZ [33] but the side information is interpolated by BID and no further process such as spatial smoothness will be applied to improve the side information;
- ii) PDWZS with BMRF side information interpolation and embedded WMCR process;
- iii) PDWZS with BID side information interpolation but without refinement;
- iv) PDWZS with BMRF side information interpolation but without refinement process.

QCIF video sequence “foreman” and “carphone” have been tested for simulation. WZ frame rate is 15 frames per second. Each pixel will be quantised with levels of $2^M=\{2,4,8,16\}$ to give four rate distortion points. The BMRF and BID interpolation is carried with block size of 16x16, full search ME window size of ± 7 pixels and the bidirectional ME window size for BID of ± 4 pixels. WMCR refinement process is carried with block size 16x16 and refinement window size of ± 2 pixels.

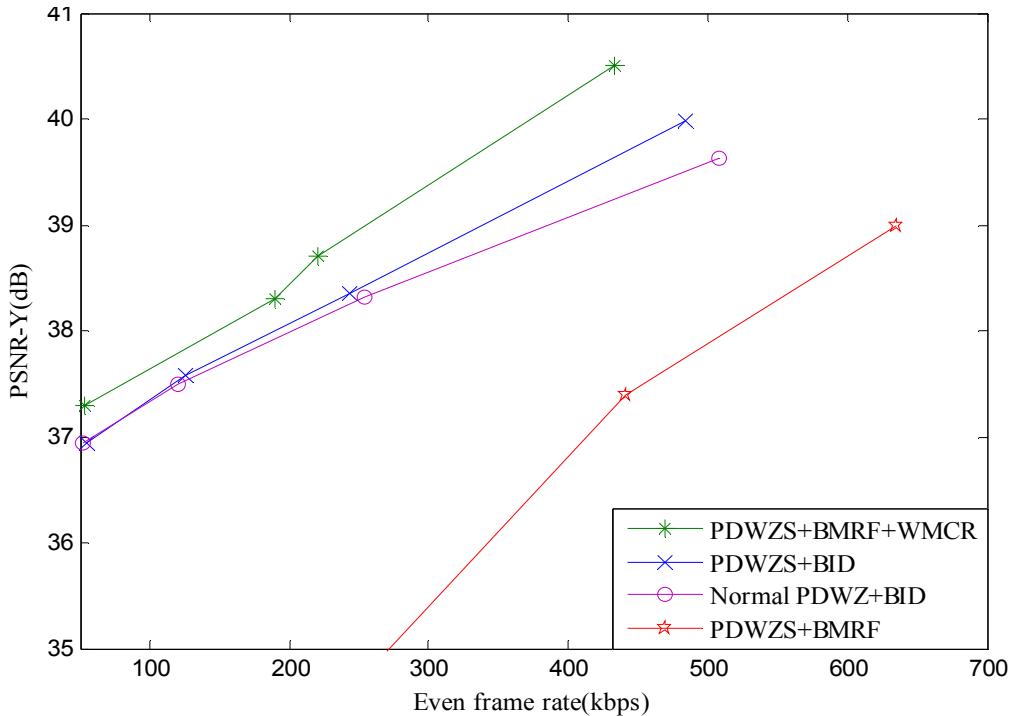


Fig.3. 14 PSNR performance of PDWZS +BMRF +WMCR refinement for Foreman

Fig. 3.14 and Fig. 3.15 show the RD performances of the proposed PDWZS with refinement. It can be observed that WZ coding with side information interpolation using BID has much better results than BMRF. This is due to the different interpolation quality of side information between these two methods. With same side information interpolation method, the PDWZS outperforms the normal PDWZ. However, we found that the performance of BMRF can be much improved by

embedding the WMCR into the turbo decoding process. The iterative process of the turbo decoder and WMCR help each other until all iteration is finished. The performance is evident in Fig.3.14 in which the PDWZS with BMRF and WMCR achieved PSNR gain of around 1 dB when compared to PDWZS with BID interpolation. It proved that PDWZS with a low complexity interpolation and a simplistic embedded refinement can achieve better system performance than PDWZ with a very complex interpolation method. In comparison with [33], the proposed model saved much computation since the WZ frame can be obtained with one round of n -iteration turbo decoding. Although embedded refinement process has slightly increased the system computation to some extent, it really counts a small portion compared to an additional whole repeated round of n -iteration turbo decoding process.

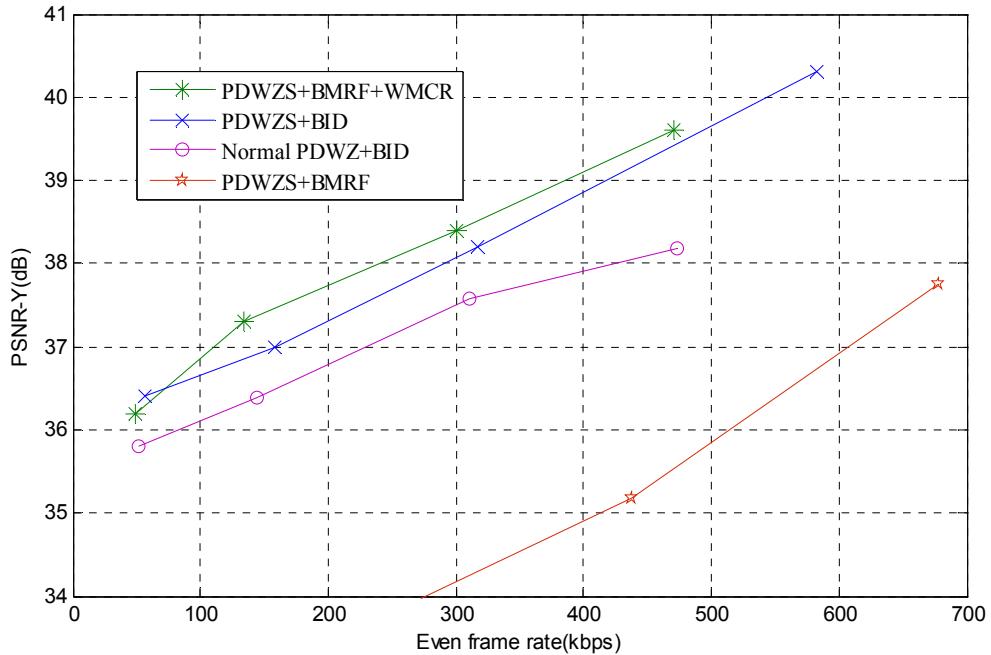


Fig.3. 15 PSNR performance of PDWZS+BMRF+WMCR refinement for Carphone

Furthermore, Table 3.3 and Table 3.4 show the decoding results and statistic change after each iteration of turbo decoder for a slice we monitored during simulation. Based on these results, it can be noticed that due to the continually improved side information, BMRF+WMCR usually can decode the slice with less iterations than purely using BID and BMRF. This brings two additional advantages of the proposed scheme. Firstly, the bit rate can be saved in the circumstance where both BID and BMRF cannot successfully decoding and more parity bits are required to send while BMRF+WMCR can decoding successfully within current limited iterations. Moreover, it is possible to design a turbo decoder with less iteration for BMRF+WMCR, this gives the benefit in the sense of system complexity since turbo iteration usually consumes much computation.

Table 3. 3 Monitored BER of embedded refinement for 4th slice of 1st WZ frame of ‘Carpone’ sequence with 16 quantisation level

Slice Bit Error Rate	BMRF	BID	BMRF+WMCR
1 st Iteration	0.0918	0.017474	0.0918
2 nd Iteration	0.0928	0.016933	0.0146
3 rd Iteration	0.0830	0.016888	0.00877
4 th Iteration	0.0854	0.016691	0
5 th Iteration	0.0933	0	0

Table 3. 4 PSNR between original 4th slice of 1st WZ frame and it’s side information and refined slice after each iteration(dB)

BMRF	BID	Refined Slice after each iteration			
		1st	2nd	3rd	4th
28.2	30.7	32.4	34.2	35.8	35.8

3.4 Adaptive feedback optimisation

In WZ video coding either in pixel domain or transform domain [25, 31], the bit rate of WZ frame is determined by the statistic dependency between the side information and the original WZ frame. Since the encoder has no knowledge of side information during encoding process, the number of how many parity bits would be enough for decoding remains unknown. Most previous literatures [3, 25, 31, 33] solve this rate control problem by relying on the process called “request-and-decode” in conjunction with rate compatible punctured turbo codec (RCPT) [27] and a feedback channel. However, the used “request-and-decode” process caused severe delay due to frequent feedback to seek the least number of parity bits to be sent. Each additional request would cause a new round of retransmission and decoding process which results huge impact on the system decoding complexity. This also makes WZ video coding hardly to be implemented in the real time applications. Moreover, the frequent retransmission at the encoder will cause more power consumption which is exactly what the WZ coding avoids.

This section introduces an adaptive feedback optimisation (AFO) scheme for PDWZS. The proposed scheme works under the model of “request-and-decode” and the request would bring back the evaluation of side information quality. As is known in WZ video coding, the quality (similarity between the side information and the original WZ frame) of side information determines the amount of parity bits to be sent. The encoder can choose the appropriate code rate based on the evaluation of side information quality and send corresponding number of parity bits. The whole decoding can be expected to be finished in very few rounds of feedback loops

therefore much system delay caused by “request-and-decode” can be greatly decreased while compression performance keeps similar.

3.4.1 Review of Rate Control

Before 2007, the rate control in DVC mainly relied on the “request-and-decode” process. From research prospective, the “request-and-decode” is very useful to investigate the performance of DVC architecture. It provides the best compression result in theory since it seeks the least number of parity bits to send to achieve required PSNR with the expense of frequent retransmission and decoding.

The “request-and-decode” process works in following way: Given the puncturing period is set to 8 for RCPT, the encoding rate choices are chosen as $(8/9, 8/10, 8/11, 8/12, \dots, 8/24)$. The encoder will first send a small subnet of parity bits which are decided by puncturing according to the highest code rate $8/9$. The decoder will perform decoding based on the received parity bits. If the decoding is failed, the decoder will send a request for more parity bits back to the encoder via feedback channel. The encoder will send another group of parity bits, with which the total number of parity bits is same as the number of parity bits after puncturing according to the next lower code rate $8/10$. This process repeats until a successful decoding result is obtained at the decoder. It can be seen that larger puncturing period of RCPT will give more coding rate choices and even smaller step of increasing the parity bits. The “request-and-decode” process almost guarantees the least number of parity bits sent thus can achieve the best compression performance in theory.

However, the expense of using “request-and-decode” algorithm is numerous retransmission, turbo decoding and huge time delay. Moreover, it is not feasible for storage application where no feedback channel is available. Thus, after 2007, many researchers started to look for solutions which can realise better rate control without using feedback channel. In [94], an rate control algorithm at the encoder is proposed. Authors assume that the encoder can access two adjacent key frames and a low complexity side information generation is performed at the encoder. With the side information at the encoder, the rate control algorithm estimates the required rate for each bit plane and determines the amount of parity bits should be sent and finally is proved to be effective after the simulation.

Similarly done in [95-97], the key frames are also decoded first at the encoder. Then coarse side information is interpolated by averaging these two key frames. The encoder will estimate the rate for each bit-plane by modelling the statistics between the coarse side information and WZ frame to be transmitted. Since the side information is less accurate than the side information

interpolated at the decoder by MCI, the rate is purposely overestimated thus to provide enough amount of parity bits. However, a clear disadvantage which can be observed for above encoder rate control algorithms is the assumption of available key frames at the encoder. This is not feasible for many DVC applications. Moreover, the ME to generate side information [94], key frame intra decoding [95, 96], and rate prediction process all increase lots of computations at the encoder, which makes the encoder not low complexity anymore and diverted from the original concept designed for DVC. In [98], a decoder rate control algorithm based on new designed BER detection and CRC is proposed.

The ideal algorithm should be encoder rate control with which the feedback channel can be eliminated. However, in order to perform the rate control at the encoder, the encoder has to have access to the side information which is hard to be implemented due to the operation will increase the computation complexity of the encoder. In some applications, the cameras are not allowed to communicate with each other where side information definitely is unavailable at the encoder. Another way to perform the rate control is decoder rate control, but the feedback channel has to be borrowed to send back the information. In fact, there exist certain DVC applications which can be compatible with the feedback channel i.e. community video surveillance system, with which the end user can tolerate small time delay of the output. Consequently, it is meaningful to investigate the decoder rate control algorithm which can optimise the feedback process. An adaptive feedback optimisation algorithm at the decoder is proposed in next section, which is targeting to optimise the “request-and-decode” process in DVC.

3.4.2 Algorithm and Operations

The proposed adaptive feedback optimisation (AFO) works for the PDWZS introduced in Section 3.1. The AFO starts to work when a slice is received and decoded unsuccessfully. Then, the side information quality will be evaluated and the evaluation will be sent back to the encoder via feedback channel to aid the encoder to make judgment on how many parity bits should be sent for next transmission. The encoder makes the judgment by utilizing the so called AFO table, which stores the mapping relationship between side information quality and proper code rate choices which can make successful decoding. The AFO table should be built beforehand by running intensive simulation with similar type of video content that is going to transmit. Once the encoder received the information of side information evaluation from feedback channel, it will look up the table according to the evaluation and find the appropriate code rate to send correct number of parity bits to the decoder. If decoding still fails, the encoder will choose next candidate code rate in the table and this process repeats until the whole decoding is finished. Comparing to the

conventional WZ coding, the decoding process can be finished in very few rounds and much system delay are expected to be saved.

3.4.2.1 Side information evaluation

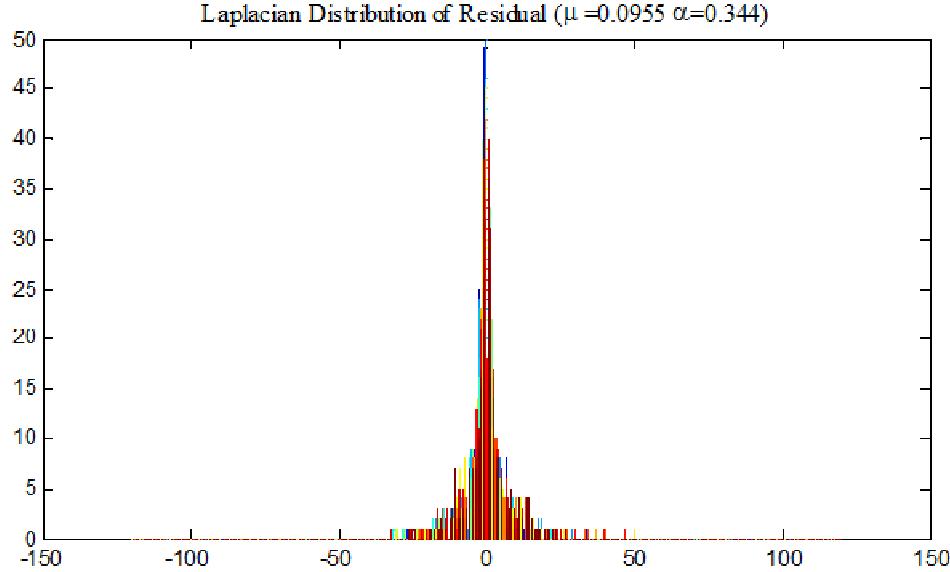


Fig.3.16 A typical residual distribution in WZ video coding

In the PDWZ video coding, an adjacent key frame is usually used to help calculate the Laplacian parameter, which is used to model the statistics between side information and the transmitted WZ frame. In the proposed AFO scheme, the variance of each slice in the side information is evaluated as the criteria to measure the slice quality.

$$\delta_s^2 = E[(r_s - E(r))^2] = E(r_s^2) - E(r)^2 = E(r_s^2) - \mu^2 \quad (3.11)$$

where r denotes the residual between side information and used key frame. μ denotes the mean value (expected value) of the residual r and δ^2 denotes the variance of the residual r . r_s denotes the residual between current corresponding slices of the side information and the key frame. The variance δ_s^2 can measure the statistical disperse degree of residual of the slices around μ . Larger δ_s^2 reflects the distance between r_s and μ , as shown in the Fig.3.16, which also implies the larger difference between slices of the side information and the key frame. In the case of $\delta_s^2 < \delta^2$, the residual r_s is more likely located close to μ thus there is higher probability to make correct prediction on the input symbol of X_{2i} with help of side information for turbo decoding. Vice versa, when $\delta_s^2 > \delta^2$, the slice of side information will be quite different from the slice transmitted, thus it is quite likely to make wrong prediction, which leads to unsuccessful turbo decoding. The

errors have to be fully corrected by receiving more parity bits from the encoder. It is observed that larger is the δ_s^2 / δ^2 , more parity bits are required. In essence, the evaluation process can be regarded as a kind of estimation of virtual channel condition.

3.4.2.2 AFO table

Based on above analysis, the AFO table is built with the purpose of deriving the relation between the range of δ_s^2 / δ^2 and the number of parity bits (namely the code rate) which can successfully decode the slice. The decoder will feedback the result of δ_s^2 / δ^2 for current slice and the encoder will look up the AFO table to find the reasonable code rate to send the corresponding number of parity bits. In order to find the target relationship between δ_s^2 / δ^2 and code rates, several simulations have to be done beforehand. The process of building the AFO table for transmitting the low motion video sequences is shown as following.

Four representing sequences “Foreman”, “Carphone”, “Coastguard” and “Silent” are picked out for simulation. For each video sequence, 140 WZ frames are encoded and transmitted. The same PDWZ with slice structure is applied, in which each WZ frame is quantised by 16 quantisation level thus each frame can be divided into 25 slices. Note 16 quantisation level is not the only choice and other options are also applicable to build the AFO table. The puncturing period of RCPT is set to 8 to provide candidate code rate range of {0,8/9,8/10,8/11...8/24}.

With above sequences, there are total 14,000 slices during simulation. For each slice, the final decoding rate and the value of δ_s^2 / δ^2 will be recorded. The AFO table is built by following ways:

- Firstly, the range of δ_s^2 / δ^2 is divided into 3 categories: $\delta_s^2 / \delta^2 < 1$, $1 \leq \delta_s^2 / \delta^2 < 2$ and $\delta_s^2 / \delta^2 \geq 2$.
- Within each category, we draw the probability distribution of each rate and determine which rates are the possible candidates that can successfully decode the slice.
- For each category of δ_s^2 / δ^2 , a list of candidate rates will be derived according to their probability histogram. The candidate code rates will be filled into AFO table with corresponding category of δ_s^2 / δ^2 .
- When encoder starts to look up the AFO table, it first locates the range of δ_s^2 / δ^2 , then find the code rate with the highest probability that can successfully decode the slice. If the chosen code rate didn't perform decoding successfully, the encoder will chose next lower code rate with second highest probability and so on.

With AFO table, the encoder can omit those rates which are not possible to perform successful decoding thus make the “request-and-decode” process quick and reliable.

Fig.3.17 shows the probability distribution $P(r)$ of each code rate in each category of δ_s^2 / δ^2 based on simulation results. The points $\{0.889, 0.8, 0.727..0.333\}$ locate at x axis is relative to the code rates of $\{8/9, 8/10, 8/11, \dots, 8/24\}$. It can be seen from Fig.3.17 (a), when $\delta_s^2 / \delta^2 < 1$, most slices can be decoded by the code rate of $8/9$. With the increase in the value of δ_s^2 , as shown in Fig.3.17 (b) and (c), the final decoding code rate decreased correspondingly.

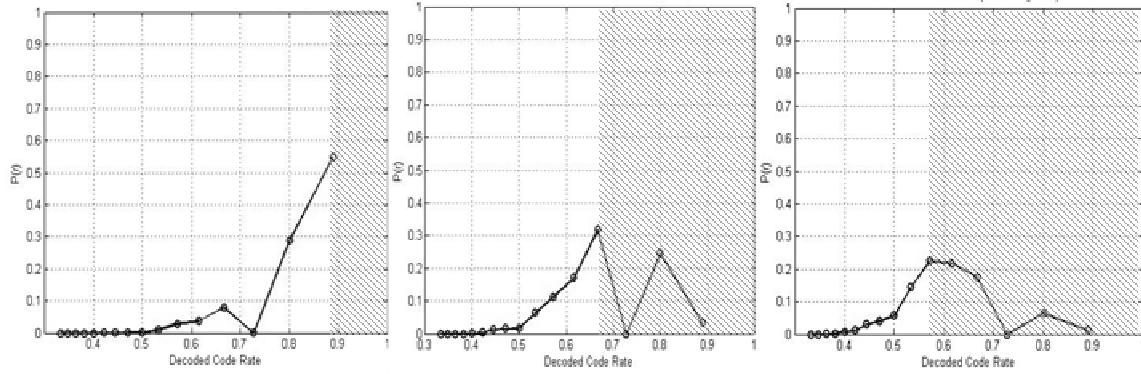


Fig.3. 17 Probability distribution of final decoding code rate with different range of

Foreman, Coastguard, Carphone, and Slient Sequences

(a) 9614 sample slices (b) 2597 sample slices (c) 1789 sample slices

Table 3.5 shows an example of AFO table built from Fig.3.17. Note the higher code rate but with lower probability (code rates locate in shadow area) and the code rate with probability of zero will not be qualified as candidate code rates. Finally, the code rate $8/24(1/3)$ will be added in case of the worst situation that all previous code rates are unable to decode successfully. Table 3.5 also includes the code rate list of ERC (Exhaustive Rate Control) and SRC (Selective Rate Control) for comparison.

Table 3. 5 Proposed AFO scheme for puncturing period of 8

Scheme	Code rate Selection list						
ERC	8/9, 8/10, 8/11, 8/12, 8/13, 8/14, 8/15, 8/16, 8/17, 8/18, 8/19, 8/20, 8/21, 8/22, 8/23, 8/24						
SRC	8/9, 8/10, 8/12, 8/16, 8/24						
AFO	<table border="1"> <tr> <td>$\delta_s^2 / \delta^2 < 100\%$</td> <td>8/9, 8/10, 8/12, 8/13, 8/14, 8/15, 8/24</td> </tr> <tr> <td>$100\% \leq \delta_s^2 / \delta^2 < 200\%$</td> <td>8/9, 8/12, 8/13, 8/14, 8/15, 8/16, 8/17, 8/18, 8/24</td> </tr> <tr> <td>$200\% < \delta_s^2 / \delta^2$</td> <td>8/9, 8/14, 8/15, 8/16, 8/17, 8/18, 8/19, 8/20, 8/24</td> </tr> </table>	$\delta_s^2 / \delta^2 < 100\%$	8/9, 8/10, 8/12, 8/13, 8/14, 8/15, 8/24	$100\% \leq \delta_s^2 / \delta^2 < 200\%$	8/9, 8/12, 8/13, 8/14, 8/15, 8/16, 8/17, 8/18, 8/24	$200\% < \delta_s^2 / \delta^2$	8/9, 8/14, 8/15, 8/16, 8/17, 8/18, 8/19, 8/20, 8/24
$\delta_s^2 / \delta^2 < 100\%$	8/9, 8/10, 8/12, 8/13, 8/14, 8/15, 8/24						
$100\% \leq \delta_s^2 / \delta^2 < 200\%$	8/9, 8/12, 8/13, 8/14, 8/15, 8/16, 8/17, 8/18, 8/24						
$200\% < \delta_s^2 / \delta^2$	8/9, 8/14, 8/15, 8/16, 8/17, 8/18, 8/19, 8/20, 8/24						

3.4.3 Results and Discussions

In order to investigate the performance of AFO scheme and make fair comparison with ERC, 5 other video sequences other than those 4 video sequences used to build up AFO tables are used to

verify the scheme. The simulation parameters are same as that of building the AFO table. To show the benefit of the AFO scheme, the number of times that the requests happen for each slice to decode successfully is recorded during simulation. Table 3.6 shows the final results of AFO scheme in comparison with ERC scheme.

Table 3.6 R-D performance comparison for AFO and ERC scheme

Video	Average Code Rate		Average request times per slice		Average PSNR(dB)		Average bitrate of WZ frame(kbps)		Delay reduction ratio $(N_{ERC}-N_{ARC})/N_{ERC}$
	ERC	AFO	N_{ERC}	N_{ARC}	ERC	AFO	ERC	AFO	
Claire	0.874	0.827	0.172	0.061	49.320	49.321	213.283	309.149	64.5%
Mobile	0.683	0.682	2.983	1.603	37.215	37.215	688.862	691.150	46.3%
Mother	0.860	0.819	0.388	0.045	45.967	45.970	240.911	327.886	88.4%
News	0.841	0.828	0.692	0.106	44.426	44.426	279.659	307.588	84.7%
Salesman	0.868	0.831	0.279	0.106	45.950	45.953	224.882	300.104	62.0%

It is evident from Table 3.6 that the R-D performance of AFO is very close to that of ERC. The average PSNR of AFO is even higher than that of ERC slightly; only the corresponding bit rates of WZ frame are slightly higher than that of ERC scheme. This is due to the AFO scheme tries to send the most probably appropriate amount of parity bits but cannot guarantee the least parity bits at each time. The advantage of AFO scheme is apparent and significant when we compare the average request times per slice. Much system delay has been eliminated due to the efficient and accurate rate control. For Mother Sequence, the proposed scheme reduced system delay up to 88.4%, and has greatly improved the system processing speed and efficiency.

To further explore the performance of AFO, we assume a practical system which can only tolerate the maximum system delay up to 3 times of feedback. Under this circumstance, each scheme would only have the first 4 code rates to be chosen during the decoding process. Fig.3.18 shows the perceptual quality comparison for the Y-component of news sequence by three rate control schemes.

It can be noticed that due to the limitation in delay tolerance, the ERC is too slow to send enough and appropriate number of parity bits therefore many slices were not decoded successfully and the worst perceptual quality is obtained. Conversely, the SRC dropped the code rate rapidly. No matter the side information quality is good or not, after 3 rounds of feedback loop, there have already been half of all parity bits are sent for decoding. Too much redundant parity bits are sent, which results in the much higher bit rate than the other two schemes. Even though the perceptual quality obtained is very high, the bandwidth is not efficiently utilised. Comparing

to ERC and SRC, the proposed AFO scheme find the correct code rate quickly based on the evaluation of side information, the whole decoding is done quickly within 3 rounds of feedback. Under the condition of low latency given, not only a satisfying perceptual quality is delivered, the bit rate is saved and the bandwidth is also efficiently utilised.

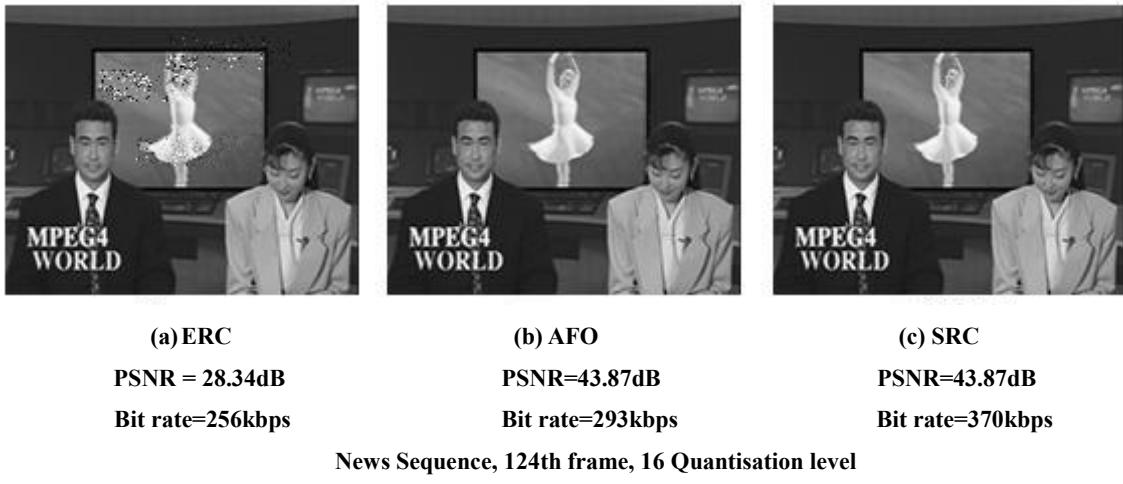


Fig.3.18 Perceptual comparison with different rate control algorithms

3.5 Chapter Summary

In this chapter, four algorithms including slice structure design, BMRF, embedded refinement and adaptive feedback optimisation are proposed to optimise the decoder complexity in WZ video coding. The complexity of the decoder has significant impact on the decoding efficiency and most current issues in WZ coding solution neglect this aspect. In fact, the decoder in WZ video can be further optimised while the same system performance still can be maintained.

First of all, a slice structure is applied in conventional WZ video coding architecture. The initial purpose is to reduce the input block size of Slepian-Wolf codec thus to fasten the overall coding process. The length of slice is determined by the 3GPP turbo codec, which limits the input block size to 5114 bits. In addition, the slice structure used can further optimise the rate control during the transmission. With the slice structure, the encoder reduces the possibility to send unnecessary parity bits during the decoding process thus the compression performance can be further improved. The simulation results proved the advantages of PDWZ with slice structure not only in the PSNR improvement, but also in reducing the decoder complexity.

To further optimise the decoder, a fast BMRF frame interpolation techniques is introduced for side information interpolation. The proposed BMRF explores the motion in multiple reference frames and performs fast interpolation. Unlike the popularly used BID MCI interpolation which is relatively complex algorithm, BMRF offers a sub optimal solution to side information generation in that it produces the side information which is much superior to that generated by normal

average interpolation or TLIN, but inferior to that generated by BID MCI. However, BMRF is much faster than BID frame interpolation. The simulation proved that BMRF saved nearly 90% computations with little expense of loss in the quality of side information.

There are many algorithms towards refine the side information thus to improve the system performance furthermore. However, this kind of operation usually operates outside the decoding process thus huge decoding complexity will be resulted in, e.g. repeat decoding. To optimise this process, an embedded refinement algorithm which moves the refinement process into the turbo decoder is proposed. The embedded refinement process operates after each iteration of turbo decoding. Inside the decoder, side information and turbo decoding are improved mutually by each other, with which turbo decoding can be done within n -iterations and no more repeat decoding is needed. Moreover, since embedded refinement process is adopted, it is not necessary to use complex side information interpolation technique, e.g. BID MCI. The introduced BMRF can be used in this case. BMRF results in the degradation of the coding performance due to the inferior quality of the side information. However this quality loss can be compensated by the embedded refinement process. In comparison with normal PDWZ with BID and refinement process outside the decoder, the proposed frame work has similar performance whilst makes the decoder more efficiency in following aspects: Firstly, the computation is saved by using simple and fast BMRF other than BID. Secondly, no repeat decoding is needed due to the embedded refinement introduced in.

Another factor cause huge system delay is the rate control algorithms adopted by most WZ video coding solutions. The rate control is performed by the “request-and-decode” process, which repeatedly requests more parity bits from the encoder until the decoding is successfully done. However, the number of requests happen during the decoding has significant impact on the complexity of decoder in that each request causes a whole round of decoding process. To reduce the number of the requests happen, an adaptive feedback optimisation algorithm is proposed. With this algorithm, the side information evaluation is performed after the decoding fails. The request containing the information of side information quality is feedback to the encoder via feedback channel. The encoder will not work “blindly” as in conventional “request-and-decode” process but choose the possible code rate according to the information feedback from the decoder. The number of requests happen can be greatly reduced with the proposed algorithm thus the complexity is reduced significantly.

To conclude this chapter, the proposed PDWZ based on slice structure, BMRF, embedded refinement and adaptive feedback optimisation has more efficient decoder than the conventional PDWZ without suffering from the performance loss. Simulation results also proved that the slice

structure can improve the RD performance to some extent. All works in this chapter is based on the PDWZ. In the next chapter, we will focus on the research of transform domain Wyner-Ziv architecture. An advanced TDWZ architecture based on the 3D DCT is proposed and its application in multiview video coding is explored.

Chapter 4

4. Transform Domain WZ Video Coding with 3D DCT and Applications

In order to further improve the WZ video coding performance and reduce its gap to that of the traditional video coding standard, an advanced TDWZ based on 3D DCT (3D-TDWZ) is proposed in this chapter. The purpose of using 3D DCT is to exploit the temporal correlation between successive WZ frames without increasing too much complexity to the encoder and achieve higher compression than current TDWZ based on 2D DCT (2D-TDWZ) meanwhile. In the later sections of this chapter, the research focuses on the Distributed Multiview Video Coding (DMVC) which is the direct application of WZ video coding. Particularly, three types of transform domain DMVC architectures based on 2D-TDWZ or 3D-TDWZ are proposed. The performance of transform domain DMVC is investigated by comparing to that of the existing pixel domain DMVC or DMVC purely using intra-frame coding.

4.1 Background and Motivation

In Chapter 2, the TDWZ is introduced which applies 2D DCT or 2D DWT to WZ frames to remove the spatial redundancy. The compression performance is improved significantly comparing to that of the PDWZ. Even though, the performance of WZ video coding still cannot match the current video coding standards which are based on the predictive coding structure. One of the major reasons is that in WZ video coding, the temporal correlation is not fully exploited. The ME and MC process usually used in predictive video coding to remove temporal redundancy are not suitable for WZ video coding due to the computation constraint. It is necessary to find an alternative choice for ME to exploit the temporal correlation in WZ video coding whilst low complexity of the encoder can be kept at same time. 3D DCT which is an extension of 2D DCT algorithm can be used in this case. 3D DCT performs another 1D DCT transformation in temporal direction apart from 2D DCT applied in spatial direction, through which the temporal correlation is exploited. The complexity is only increased by additional temporal transformation which is much lower than that caused by ME at the encoder. In this chapter, a new TDWZ architecture based on 3D-DCT transformation to exploit the correlation of both spatial and temporal direction of WZ frames is introduced. With 3D-TDWZ, the performance of WZ video coding can be further enhanced.

One of the most important applications of WZ video coding is the Multiview video coding

which is used to deal with multiple video contents collected from multiple cameras. Multiview video can provide more vivid and accurate video information than the traditional monoview video and can be applied in some future emerging applications, e.g. 3D TV, multiple cameras system, free viewpoint switching. However, the amount of video data to be processed is very large. To efficiently store or transmit these data, effective compression algorithms have to be designed. There are numerous multiview video coding algorithms and many of them are based on the interview prediction which requires cameras to be able to communicate with each other and camera has to be equipped with high processing ability. These conditions are unlikely to be realised in many practical applications where cameras deployed in the scene are energy constrained with low complexity. Recently, distributed multiview video coding (DMVC) based on WZ video coding is proposed in [99-104] to offer the solution for such multiview case that has constraint for the encoder. In DMVC, each camera is separately encoded and jointly decoded. With WZ video coding, each camera keeps low complexity and achieves compression efficiency via joint decoding. Many current DMVC works are based on pixel domain, in which cameras apply PDWZ to encode the frames. Since TDWZ provides superior performance to PDWZ, it is reasonable to design a transform domain DMVC via applying TDWZ other than PDWZ to compress the frames.

4.2 3D DCT

In current major video coding standards, 2D DCT transformation is widely used to remove the spatial redundancy while the temporal redundancy is removed by motion estimation and prediction process. But ME has following disadvantages:

- ME may not be able to yield good motion vectors in certain type of high motion video sequences.
- ME cannot perform well for non-translational motions, such as rotation, zooming, and shearing of objects.
- ME inevitably causes huge complexity at the encoder, while in some applications the encoder is energy constrained so that cannot afford to perform ME.

In order to avoid the problems caused by ME, 3D DCT can be used in video compression as an alternative choice to exploit both spatial and temporal correlation. Since the digital video sequence can be conceived as 3-dimensional signal, it is natural to extend 2D DCT algorithm to 3D DCT by performing 3rd 1D DCT in temporal direction. The algorithm of applying 3D DCT in video encoding typically involves following processes:

- Stacking a group of frames.

- Dividing frames into pixel cubes
- Applying 3D DCT transformation to every pixel cube
- Quantisation of coefficients
- Scanning and entropy coding

The forward 3D-DCT can be given by

$$Y_{u,v,w} = \sqrt{\frac{8}{MNP}} E_u E_v E_w \times \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{p=0}^{P-1} y_{m,n,p} C_{2M}^{(2m+1)u} C_{2N}^{(2n+1)v} C_{2P}^{(2p+1)w} \quad (4.1)$$

where

$$E_u, E_v, E_w = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } m, n, p = 0 \\ 1 & \text{otherwise} \end{cases}$$

$$C_J^i = \cos\left(\frac{i\pi}{J}\right)$$

$y_{m,n,p}$ is the 3D spatio-temporal data element of the m th row, n th column and p th frame;

$Y_{u,v,w}$ is the 3D transform domain data element at position u, v, w in the 3D transform space;

M, N, P are the dimensions of the data cube.

Furthermore, $Y_{u,v,w}$ can be normalised as

$$\bar{Y}_{u,v,w} = \sqrt{\frac{MNP}{8}} \frac{1}{E_u E_v E_w} Y_{u,v,w} \quad (4.2)$$

Applying equation (4.2) into equation (4.1), then

$$\bar{Y}_{u,v,w} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{p=0}^{P-1} y_{m,n,p} C_{2M}^{(2m+1)u} C_{2N}^{(2n+1)v} C_{2P}^{(2p+1)w} \quad (4.3)$$

The inverse 3D-DCT (IDCT) for the spatio-temporal data points can be represented by

$$y_{m,n,p} = \sqrt{\frac{8}{MNP}} \times \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \sum_{w=0}^{P-1} E_u E_v E_w \bar{Y}_{u,v,w} C_{2M}^{(2m+1)u} C_{2N}^{(2n+1)v} C_{2P}^{(2p+1)w} \quad (4.4)$$

Furthermore, $\bar{Y}_{u,v,w}$ can be normalised as

$$\bar{Y}_{u,v,w} = \sqrt{\frac{8}{MNP}} E_u E_v E_w Y_{u,v,w} \quad (4.5)$$

With equation (4.4) and equation (4.5), we have

$$y_{m,n,p} = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \sum_{w=0}^{P-1} \bar{Y}_{u,v,w} C_{2M}^{(2m+1)u} C_{2N}^{(2n+1)v} C_{2P}^{(2p+1)w} \quad (4.6)$$

3D DCT actually is used to compress motion pictures early in 1970s [105, 106], where significant improvement in compression performance resulting in higher compression than

conventional intra-frame coding is reported. Afterwards, there have been many works done to improve the video codec based on 3D DCT. In [107], 3D DCT is used in video compression where different coding algorithms are applied to the coefficients belonging to different ranges of values,. In [108, 109], authors mainly contribute to how to build quantisation volume (3D quantisation matrix) and set up efficient scan order for the coefficients. In [110], the hardware development of video codec adopting 3D DCT is mainly discussed. In[111], the advantages and disadvantages of 3D DCT video coding are discussed. In [112], an adaptive 3D-DCT compression technique is proposed, which dynamically determines the optimal size of the video cube based on the motion analysis inside. Similar work has been done in [113], however, the motion analysis is performed in DCT domain. In order to reduce the computation complexity of 3D DCT, in [114, 115], fast 3D DCT transformation algorithms are proposed. In [116, 117], the applications of 3D DCT in multiview video coding are discussed.

Comparing to ME, 3D DCT has following advantages:

- It avoids the problems that ME causes.
- It causes a symmetric codec.
- It has much simpler algorithm. 3D DCT can be considered as a series of 1D DCT transformation in different direction
- It makes a low complexity encoder.
- It produces higher reconstruction quality frame.

However, applying 3D DCT inevitably causes following problems:

- It requires larger memory to store data. Larger pixel cube will increase the memory requirement correspondingly.
- Comparing to the predictive coding with GOP structure of IPPPI structure, the system output is not frame by frame but group by group.
- The higher quality output is compensated by the moderate loss in compression ration comparing to predictive coding.

Although the application of 3D DCT in video compression has several drawbacks, it is much more suitable than ME for the WZ video coding. It satisfies the requirement of low complexity encoder. It also exploits the temporal correlation thus to provide higher compression performance. For the purpose of building an advanced TDWZ, applying 3D DCT in TDWZ is worthy of studying.

4.3 Transform Domain WZ Video Coding with 3D DCT

4.3.1 Architecture and Operations

In this section, a 3D DCT based TDWZ (3D-TDWZ) is proposed. Fig.4.1 illustrates the proposed architecture which is similar as the architecture of previous TDWZ proposed in [3, 31]. The main difference is that the transformation used here is 3D DCT and WZ frame is not encoded frame by frame but encoded group by group.

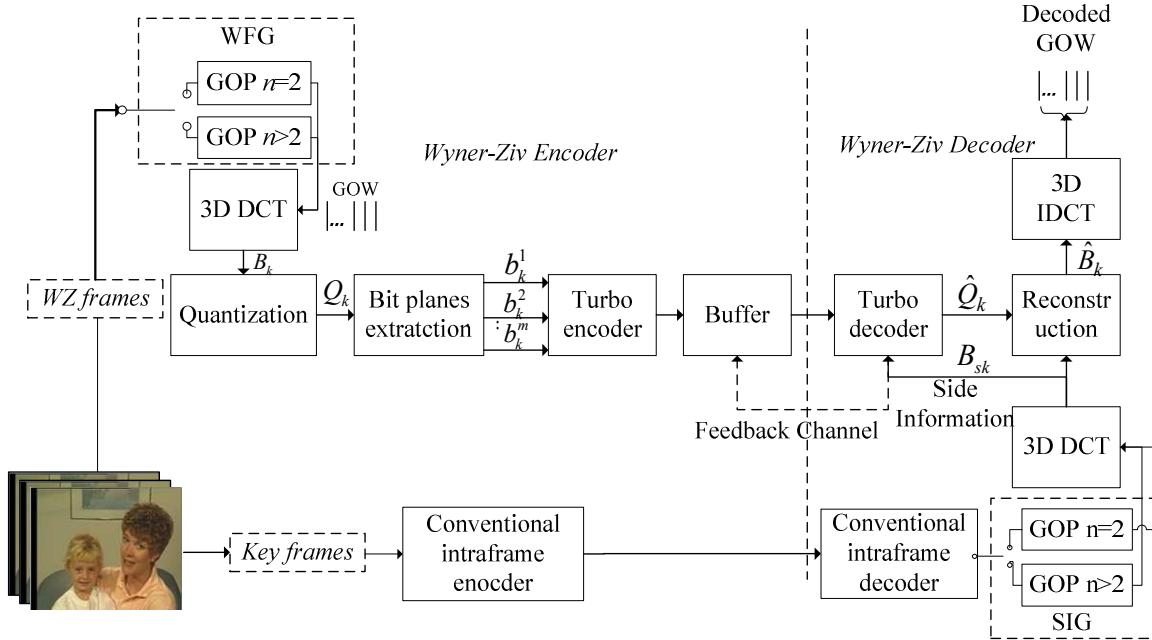


Fig.4. 1 The architecture of transform domain DVC with 3D DCT

The operation of the 3D-TDWZ is explained by following steps.

- The input video sequence is divided into key frame sequence and WZ frame sequence for transmission. Key frames are coded via conventional intra-frame coding. WZ frames are inter-frame transform coded with 3D DCT at the encoder and inter-frame decoded at the decoder with the help of side information.
- Sequence decomposition of WZ frames is performed according to the length of the GOP. Given the GOP length is n , the first frame in a GOP is key frame and the rest are WZ frames. The length of GOP determines how many WZ frames lie between two adjacent key frames. For example, $n=2$ is the most common case in WZ video coding in which there is only one WZ frame between two key frames. Then key frame sequence is obtained by grouping all odd frames and WZ frame sequence is obtained by grouping all even frames. For $n>2$, there are more than one WZ frames between two key frames.
- In order to apply 3D DCT, the WZ frame sequence is organised into groups which is called Group of WZ frames (GOW) for process. This process is done in the unit of WZ

frame grouping (WFG). Note that the operation to organise the GOW is different for GOP=2 and GOP>2. The details can be referred to later section regarding to WFG.

- For each GOW, a $4 \times 4 \times l$ Cubewise 3D DCT then is applied to generate several DCT cubes, where l is the length of GOW.
- Coefficients at the same position in every DCT cube of the GOW will be grouped together to compose the coefficient bands. The total number of coefficient bands are $4 \times 4 \times l$.
- For each coefficient band B_k where k is the index of the coefficient band, the coefficients are quantised using a uniform scalar quantiser with 2^{M_k} levels. The quantisation level for each coefficient band is determined by a predefined $4 \times 4 \times l$ quantisation volume (3D quantisation matrix). The detail process of deriving quantisation volume is described in later section.
- After the quantisation, the quantised symbols Q_k are converted into fixed-length binary codewords and corresponding bit-planes extraction is performed. The bits with same significance of each quantised symbol are blocked together to form M bit-planes.
- Each bit-plane b_k^j ($1 \leq j \leq M$) is sent to the Slepian-Wolf encoder for encoding in the order from the most significant bit-plane to the least significant bit-plane.
- Here a rate-compatible punctured turbo code (RCPT) is implemented as Slepian-Wolf codec in combination with a feedback channel. The feedback channel is used to adapt the statistical change between the current band and the corresponding band of side information.
- After encoding, the parity bits produced by the turbo encoder are stored in a buffer and only a subset of them is transmitted to the decoder upon request.
- At the decoder, corresponding side information for the current GOW are generated in the unit of side information grouping (SIG) by using received key frames.
- The generated side information frames will be grouped and transformed by same cubewise 3D DCT as that performed at the encoder and the coefficient bands are blocked together. The corresponding coefficient band B_{sk} then is used to help turbo decoder decode the transmitted bit-plane.
- To help decoding, the correlation between the coefficient bands B_k and B_{sk} is modelled by Laplacian distribution. The corresponding Laplacian parameter is calculated offline, with which and B_{sk} the soft input of the turbo decoder is predicted.
- After decoding, if the BER of the decoded bit plane is higher than a predefined threshold, e.g. 10^{-3} , the decoder will request more parity bits from the encoder and whole decoding process will repeat again until the current bit-plane is successfully received.

- All decoded bit-planes are grouped together and the quantised symbols \hat{Q}_k are reconstructed.
- Combining \hat{Q}_k with the corresponding coefficient band B_{sk} from side information, the current coefficient band \hat{B}_k will be reconstructed as $E(\hat{B}_k | \hat{Q}_k, B_{sk})$. The purpose of this operation is to bind the magnitude of the reconstruction distortion to a maximum value, which is determined by the quantiser coarseness.
- After all coefficients bands are grouped, inverse 3D DCT is performed and the GOW is reconstructed. The next GOW will start to transmit.

4.3.2 WZ Frame Grouping (WFG)

To exploit the temporal correlation among the successive WZ frames using 3D DCT, it is necessary to organise the WZ frame sequence to groups (GOW) for processing. The function of the unit WFG is to divide WZ frame sequence into GOWs with length l . larger value of l leads to larger memory requirement at the encoder. The way to organise the GOW will be different for GOP length $n=2$ and $n>2$ and the details of composing the GOW are given as follows.

I) $n=2$

This is the most popular case used in the research. In this case, there is only one key frame and one WZ frame in a GOP. The process is shown in Fig.4.2. It should be noticed that for this type of composition is not suggested for the video sequence with low correlation.

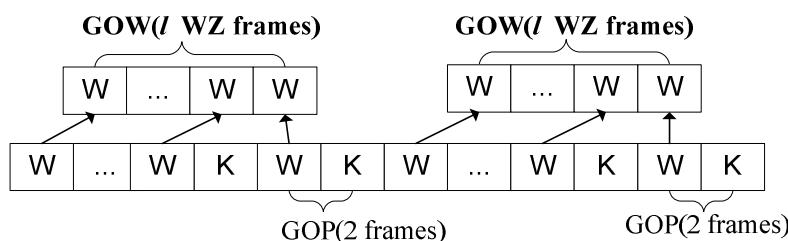


Fig.4. 2 WFG for $n=2$

II) $n>2$

When $n>2$, there are more than one WZ frames in a GOP. The GOW can be composed of all rest WZ frames of the GOP. In this case, the GOW length $l=n-1$ as shown in Fig.4.3.

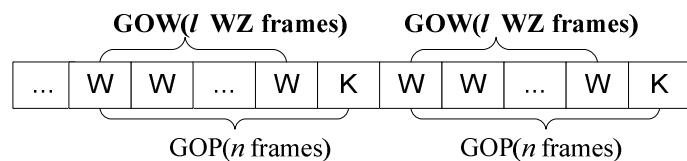


Fig.4. 3 WFG for $n>2$ and $l=n-1$

Note that l is not necessary to be $n-1$ but can be bigger than or equal to n which results in the WZ frames consisted in the GOW are located in more GOPs, as shown in Fig.4.4. However, this kind of composition can lead to further system output delay because the GOW only can be reconstructed until all the key frames involved in side information generation are received at the decoder.

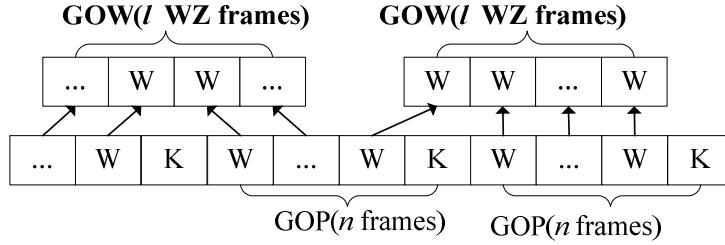


Fig.4. 4 WFG for $n>2$ and $l=n$

4.3.3 Side Information Grouping (SIG)

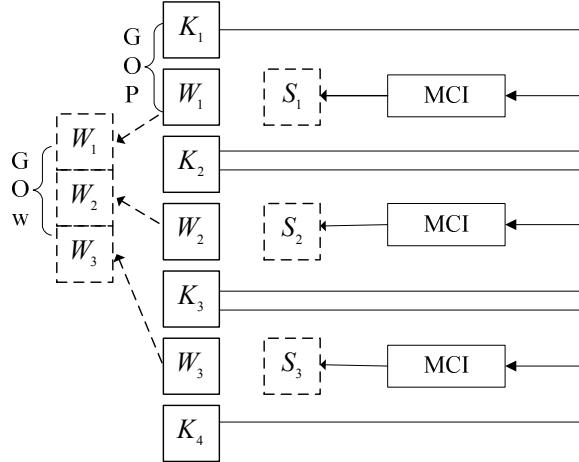
In the proposed 3D-TDWZ, SIG is used to generate corresponding side information for WZ frames in a GOW by using received key frames. Based on the assumption of linear translation model, the side information generation is implemented by performing MCI between the selected key frames. Since the GOW composition is different according to the GOP length n and GOW length l , the corresponding side information generation methods also varies. The essence of side information generation is how to interpolate one or more side information frames between two key frames. The shorter is the distance between two key frames, higher quality is the side information generated. Vice versa, larger is the GOP size and GOW size, longer is the distance between two key frames, and lower is the quality of side information. However, overall compression ratio can be improved in this case since more frames are coded by WZ coding other than intra-frame coding. The side information generation method with different lengths of GOP is mainly discussed in [26] and described as follows.

Let $MCI(A, B, d)$ represent the result of MCI function between frames A and B at d fractional distance from frame A . The side information generation methods for different GOW composition method discussed in Section 4.3.2 are presented. In the following part, $K_1K_2\dots$ denote the key frames lie in adjacent GOPs and $W_1W_2W_3\dots$ are WZ frames in the GOW and $S_1S_2S_3\dots$ are corresponding generated side information of WZ frames $W_1W_2W_3\dots$.

I) $n=2$

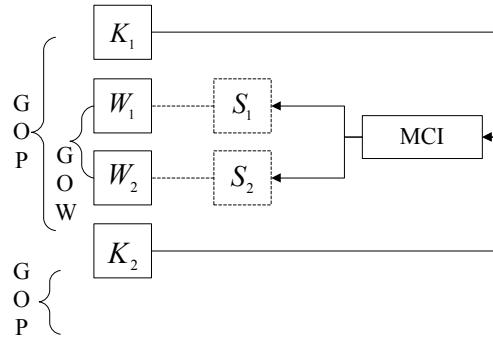
In this case, side information of each WZ frame in the GOW is interpolated by performing MCI between two adjacent key frames. Fig.4.5 shows the side information generation process for GOW containing 3 WZ frames.

$$\begin{cases} S_1 = MCI(K_1, K_2, 1/2) \\ S_2 = MCI(K_2, K_3, 1/2) \\ S_3 = MCI(K_3, K_4, 1/2) \end{cases} \quad (4.7)$$

Fig.4. 5 Side information Interpolation for $n=2, l=3$ II) $n>2, l=n-1$ (i) $n=3, l=2$

As shown in Fig.4.6, there are two WZ frames in a GOP. The corresponding side information \$S_1\$ and \$S_2\$ can be obtained by

$$\begin{cases} S_1 = MCI(K_1, K_2, 1/3) \\ S_2 = MCI(K_1, K_2, 2/3) \end{cases} \quad (4.8)$$

Fig.4. 6 Side information interpolation for $n=3, l=2$ (ii) $n=4, l=3$

As shown in Fig.4.7, there are 3 WZ frames in a GOP. The corresponding side information \$S_1\$, \$S_2\$, and \$S_3\$ can be obtained via equation (4.9):

$$\begin{cases} S_2 = MCI(K_1, K_2, 1/2) \\ S_1 = MCI(K_1, S_2, 1/2) \\ S_3 = MCI(S_2, K_2, 1/2) \end{cases} \quad (4.9)$$

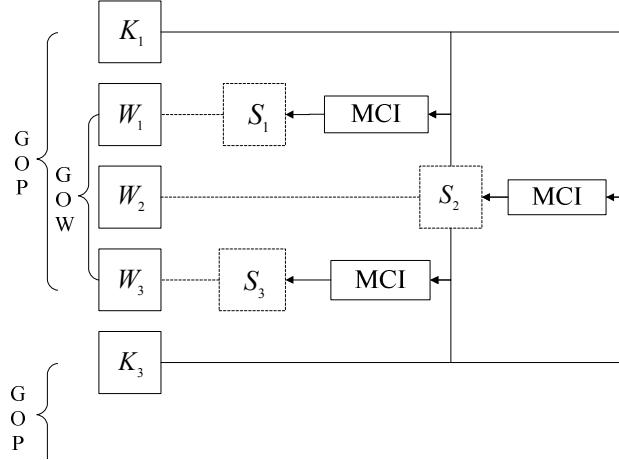


Fig.4. 7 Side information interpolation with $n=4, l=3$

With above procedures, the side information generation method for the case that GOP length $n > 4$ and GOW length $l = n - 1$ can be derived by repeating above procedures.

III) $n > 2, l \geq n$

For the case $n > 2$ and $l \geq n$, side information can be obtained by mixing methods introduced in sub-sections I) and II). Fig.4.8 shows an example of side information generation for the case $n=3$ and $l=3$.

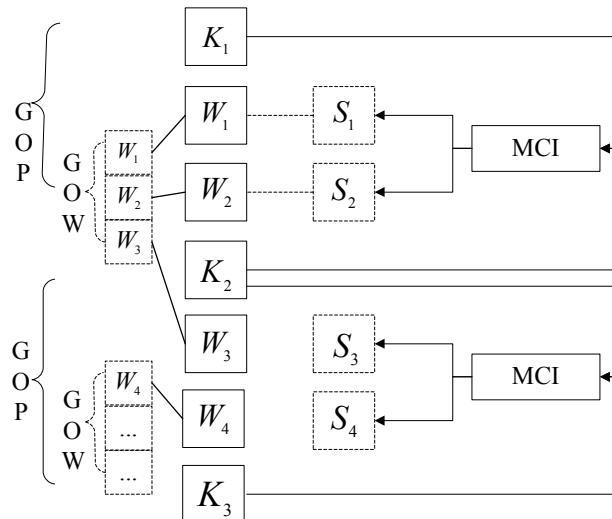


Fig.4. 8 Side information interpolation for $n > 2, l \geq n$

From Fig. 4.8, S_1, S_2, S_3, S_4 are defined as follows:

$$\begin{cases} S_1 = MCI(K_1, K_2, 1/3) \\ S_2 = MCI(K_1, K_2, 2/3) \\ S_3 = MCI(K_2, K_3, 1/3) \\ S_4 = MCI(K_2, K_3, 2/3) \end{cases} \quad (4.10)$$

4.3.4 Quantisation Volume Design

In order to investigate the RD performance of the proposed architecture, we need to setup a quantisation volume (3D quantisation matrix) to determine the quantisation levels for each coefficient band. In most 2D DCT based coding techniques, such as JPEG and MPEG, a quantisation matrix is usually used. However, it is not directly applicable for the situation with 3D DCT based coding since it cannot cope with variables in temporal axis. There have been many references regarding to how to construct quantisation matrix for 2D DCT, but few works so far have been done regarding to construct an effective 3D quantisation volume for 3D DCT. In 3D DCT transformation, It is observed that the dominant coefficients (significant coefficients including DC and part of AC coefficients) are spread along the major axes (x, y, z) of the coefficients cube, as shown in the shadow area in Fig.4.9. The dominant coefficients typically contain a high percentage of the total energy.

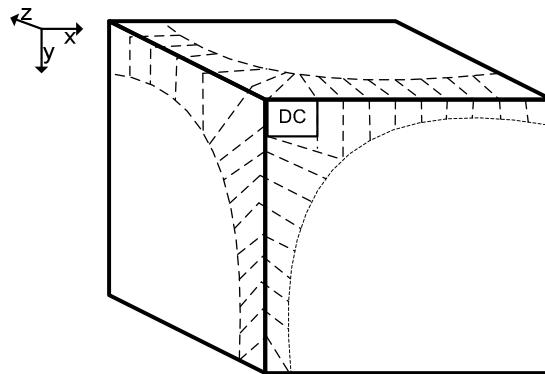


Fig.4. 9 Distribution of dominant coefficients in DCT coefficient cube

Let $C(u,v,w)$ denote the coefficient located on (u,v,w) in the cube where u, v and w are the coordinates on three major axes (x, y, z). It can be found that coefficients located farther to the DC($0,0,0$) tend to be less significant and should be more coarsely quantised. Based on these observations, a technique to build up quantisation volume for 3D DCT is proposed in [109]. Firstly, the distribution of dominant coefficients is shaped as a complement shifted hyperboloid, as shown in Fig.4.10[109], which can be described by a function $f(u,v,w) \leq C$. Here, C is an arbitrary constant to determine the size of shape and $f(u,v,w) = u^*v^*w$.

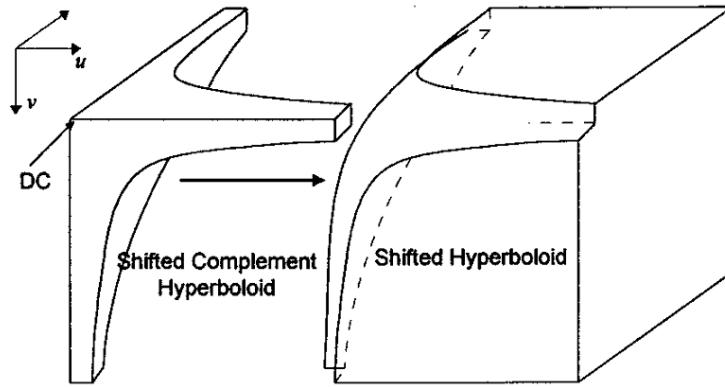


Fig.4. 10 Shape of the distribution of the dominant coefficients[109]

Then, the quantisation volume containing quantisation values for each coefficient band is build up as follows:

$$q(u, v, w) = \begin{cases} A_i(1 - \frac{e^{-\beta_i(u+1)(v+1)(w+1)}}{e^{-\beta_i}}) + 1; & \text{for } f(u, v, w) \leq C \\ A_o(1 - e^{-\beta_o(u+1)(v+1)(w+1)}); & \text{for } f(u, v, w) > C \end{cases} \quad (4.11)$$

where $q(u, v, w)$ is the quantisation value for coefficient $C(u, v, w)$, A_i and A_o are the initial amplitude respectively, which are set to the maximum quantisation value. β_i and β_o denote the decay speed to the A_i and A_o , which make quantisation value tend to be smaller inside the selected shape and bigger outside the shape. This function satisfies the requirement of quantising those dominant coefficients located inside the shape with little distortion, guaranteeing the better quality of reconstructed frame, and quantising high frequency coefficients outside the shape to zero. Parameters A_i, A_o, C, β_i and β_o can be determined experimentally and the quantisation volume can be build up afterwards.

During the simulation, the maximum of 8 bits (256 levels) is used to quantise the DCT coefficients. In order to do that, the coefficients are required to be scaled down as following:

$$S_q = \left\lceil \frac{\max(C(u, v, w))}{255} \right\rceil \quad (4.12)$$

where the numerator represents the maximum value of DCT coefficients, $\lceil x \rceil$ denotes the smallest integer bigger than or equal to x and S_q denotes the scaling factor. Next step is to build up the quantisation volume. A_i and A_o are set to 255 since the maximum value after scaling down is 255. To determine the optimal range of C, β_i and β_o , we investigate several sequences first and test the parameter influence by fixing two parameters and varies another one. With the simulation results, it is founded that parameter C has the most direct influence to the coarseness of the quantisation

Table 4. 1 An example quantisation volume($C=15$, $\beta_i=0.001$, $\beta_o=0.03$)

W=0								
U\V	0	1	2	3	4	5	6	7
0	1	2	2	2	2	2	2	2
1	2	2	2	2	2	2	2	2
2	2	2	2	2	2	120	131	
3	2	2	2	2	131	145	158	
4	2	2	2	135	152	166	179	
5	2	2	131	152	169	183	195	
6	2	120	145	166	183	197	208	
7	2	131	158	179	195	208	218	
W=1								
U\V	0	1	2	3	4	5	6	7
0	2	2	2	2	2	2	2	2
1	2	2	2	2	2	131	145	158
2	2	2	2	131	152	169	183	195
3	2	2	131	158	179	195	208	218
4	2	2	152	179	199	213	224	232
5	2	131	169	195	213	226	235	241
6	2	145	183	208	224	235	242	247
7	2	158	195	218	232	241	247	250
W=2								
U\V	0	1	2	3	4	5	6	7
0	2	2	2	2	2	120	131	
1	2	2	2	131	152	169	183	195
2	2	2	142	169	189	205	217	226
3	2	131	169	195	213	226	235	241
4	2	152	189	213	229	238	245	249
5	2	169	205	226	238	246	250	252
6	120	183	217	235	245	250	252	254
7	131	195	226	241	249	252	254	255
W=3								
U\V	0	1	2	3	4	5	6	7
0	2	2	2	2	2	131	145	158
1	2	2	131	158	179	195	208	218
2	2	131	169	195	213	226	235	241
3	2	158	195	218	232	241	247	250
4	2	179	213	232	243	249	252	253
5	131	195	226	241	249	252	254	255
6	145	208	235	247	252	254	255	255
7	158	218	241	250	253	255	255	255
W=4								
U\V	0	1	2	3	4	5	6	7
0	2	2	2	2	135	152	166	179
1	2	2	152	179	199	213	224	232
2	2	152	189	213	229	238	245	249
3	2	179	213	232	243	249	252	253
4	135	199	229	243	250	253	254	255
5	152	213	238	249	253	254	255	255
6	166	224	245	252	254	255	255	255
7	179	232	249	253	255	255	255	255
W=5								
U\V	0	1	2	3	4	5	6	7
0	2	2	2	131	152	169	183	195
1	2	131	169	195	213	226	235	241
2	2	169	205	226	238	246	250	252
3	131	195	226	241	249	252	254	255
4	152	213	238	249	253	254	255	255
5	169	226	246	252	254	255	255	255
6	183	235	250	254	255	255	255	255
7	195	241	252	255	255	255	255	255
W=6								
U\V	0	1	2	3	4	5	6	7
0	2	2	120	145	166	183	197	208
1	2	145	183	208	224	235	242	247
2	120	183	217	235	245	250	252	254
3	145	208	235	247	252	254	255	255
4	166	224	245	252	254	255	255	255
5	183	235	250	254	255	255	255	255
6	197	242	252	255	255	255	255	255
7	208	247	254	255	255	255	255	255
W=7								
U\V	0	1	2	3	4	5	6	7
0	2	2	131	158	179	195	208	218
1	2	158	195	218	232	241	247	250
2	131	195	226	241	249	252	254	255
3	158	218	241	250	253	255	255	255
4	179	232	249	253	255	255	255	255
5	195	241	252	255	255	255	255	255
6	208	247	254	255	255	255	255	255
7	218	250	255	255	255	255	255	255

volume. With this results ,we set $\beta_i = 0.001$ and $\beta_o = 0.03$ and vary the value of C to generate different quantisation volumes, which gives different rate distortion points. Smaller C denotes coarser quantisation and vice versa. With one quantisation volume, we can quantise the

coefficient $C(u,v,w)$ as:

$$Q(u,v,w) = \text{round}\left(\frac{|C(u,v,w)|}{S_q * q(u,v,w)}\right) \quad (4.13)$$

$$Q_l = \left\lceil \frac{255}{q(u,v,w)} \right\rceil \quad (4.14)$$

where $q(u,v,w)$ corresponds to the quantisation step defined in quantisation volume and $Q(u,v,w)$ is the quantised symbol of coefficient $C(u,v,w)$ and Q_l is the quantisation level needed. $\lfloor x \rfloor$ denotes the biggest integer smaller than or equal to x . The $Q(u,v,w)$ will be converted into binary symbol with $\lceil \log_2 Q_l \rceil$ bits. For the coefficients with $Q < 4$, we do not quantise it but use side information to replace it. Table 4.1 shows an example of $8 \times 8 \times 8$ quantisation volume with $C=15$, $\beta_i=0.001$ and $\beta_o=0.03$ which shows the quantisation step for each coefficient band.

4.3.5 Results and Discussions

In the simulation, turbo code used is composed by two identical constituent convolutional encoder of rate 1/2 with constraint length of 4 and polynomial generator of $g=(13, 11)$. Together with puncturing pattern, the rate compatible punctured turbo (RCPT) codec is employed as in [27] to provide flexible rates to send different sizes of subnets of parity bits to decoder. The puncturing period is set to 8 to provide the coding rate of $(0,8/9,8/10,8/11,8/12,8/13\dots 1/3)$ where coding rate 0 means no parity bits are sent but only side information are used for decoding. The simulation results are presented in two categories according to the GOP length n : $n=2$ and $n>2$, in which different simulation parameters are applied.

I) $n=2$

In the simulation, QCIF video sequences “Mother and daughter” with frame rate of 30fps are investigated. The GOW length is $l=4$ and WZ frame rate is 15fps. The first 97 frames are simulated, which consists of exact 12 GOWs. For each GOW, parameter C used for building up quantisation volume varies in the range of (1, 3, 5, 7), and $\beta_i = 0.001$ and $\beta_o = 0.03$. Bidirectional MCI is used to generate side information in SIG. In order to compare the performance, the results of 2D-TDWZ from STD research group [31] and IST research group [28] are presented as reference. It should be note that in [28, 31] 50 WZ frames are investigated, while in our simulation, the number of WZ frames to be investigated is 48. H.263+ inter-frame coding with I-P-I-P structure and DCT based intra frame coding are also used for comparison. Note in this simulation, we only present the PSNR of Y component of WZ frames since other frames (key frame or I frame) are intra-coded and has same compression result to all algorithms.

In Fig.4.11, It can be clearly observed that the proposed 3D-TDWZ has superior performance

to previous available 2D-TDWZs. For “Mother and Daughter” sequence, the PSNR gain achieves around 3.9dB more than that of STD TDWZ solution. There is maximum 1.8dB difference between IST TDWZ and 3D-TDWZ. 3D-TDWZ even outperforms H.263+ inter-frame coding at lower bit rate. However, in this case the GOP length $n=2$ which means half number of the frames are intra-coded thus the overall compression ratio is low. To improve the compression performance, larger GOP ($n>2$) should be considered.

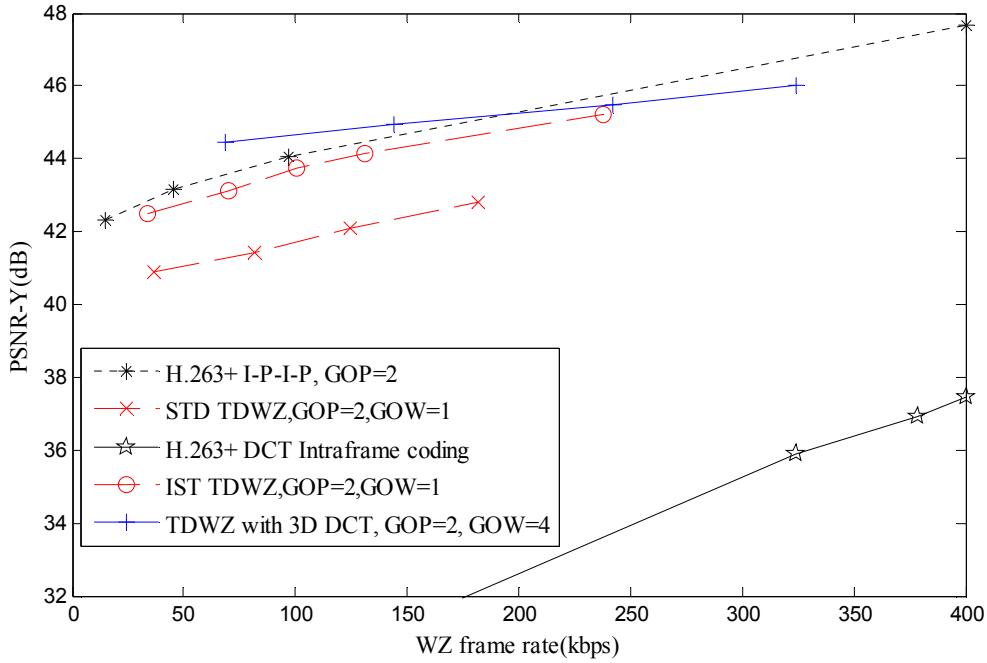


Fig.4. 11 R-D performance of 3D-TDWZ with GOP length $n=2$ of Mother and daughter sequence

II) $n>2$

In this simulation, “Salesman” and “Hall monitor” are tested. For each sequence, the first 101 frames are coded and WZ frame rate is 10fps. The proposed 3D-TDWZ are investigated with different lengths of GOP ($n=4,8$). In addition, previous available 2D-TDWZ from STD group [3], H.263+ inter-frame coding and DCT based intra frame coding are used as references for comparison. We only focus on the case that GOW length $l=n-1$. For each GOW, parameter C used for building up quantisation volume varies in the range of (1,3,5,7) for $n=4$ and (2, 4, 8, 12) for $n=8$ to obtain 4 RD points of each case. $\beta_i = 0.001$ and $\beta_o = 0.03$. Bidirectional MCI is used for side information generation.

Fig.4.12 and Fig.4.13 illustrate the performance of 3D-TDWZ with a large GOP. The similar trend as shown in Fig.4.11 can be observed that 3D-TDWZ generally outperforms the 2D-TDWZ with same GOP size and H.263+ intra-frame coding. At lower bit rate, 3D-TDWZ even outperforms the H.263+ inter-frame coding with same GOP size. Take “Hall” sequence with

GOP=4 as an example, there is around 5.1dB improvement in PSNR of 3D-TDWZ comparing to that of STD 2D-TDWZ. For the bit rate lower than 200kbps, the PSNR of TDWZ with 3D DCT is higher than that of H.263+ inter-frame coding.

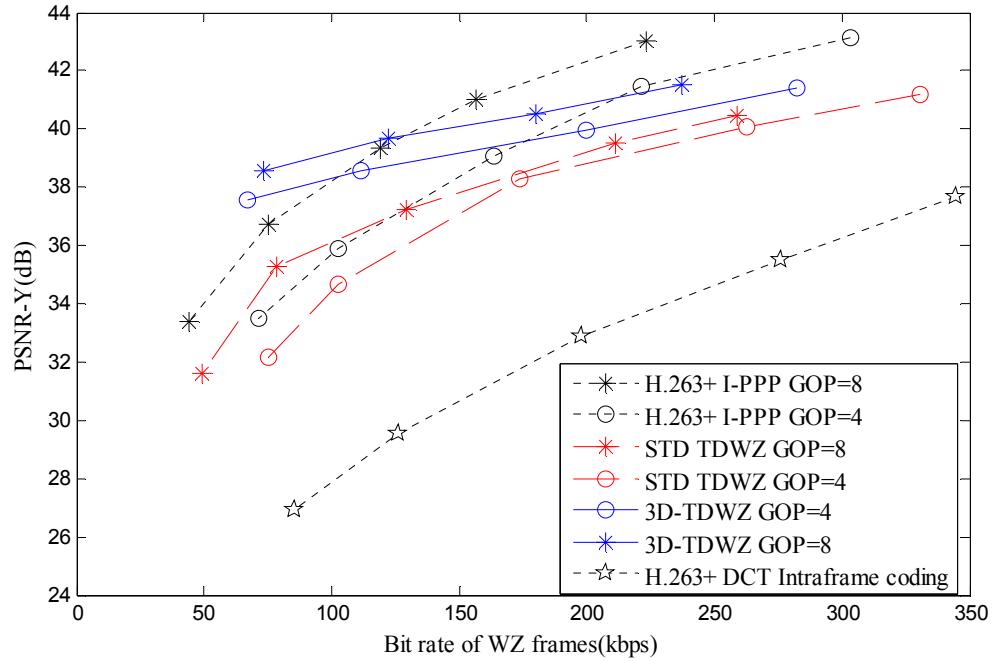


Fig.4. 12 R-D performance comparison for 3D-TDWZ for Hall sequence

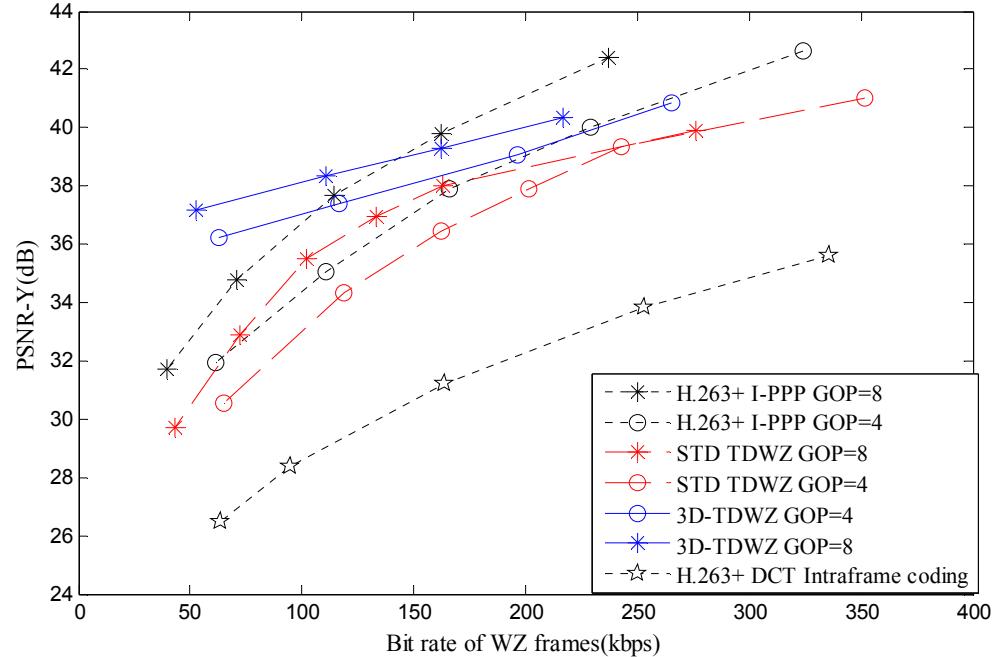


Fig.4. 13 R-D performance comparison for 3D-TDWZ for Salesman sequence

It also can be observed that larger size of GOP leads to higher overall compression. Note this result is the average PSNR performance of all WZ frames and key frames. With larger size of

GOP, the number of key frames which are coded in the intra way is greatly diminished. Thus the number of bits consumed to encode the key frame is also largely reduced. More frames are coded in WZ fashion result the number of bits to encode the WZ frames increased. The average number of bits used for WZ coding for a frame is generally much less than that used for intra-frame coding, thus the overall compression ratio is improved with the increase in the size of GOP. This trend is evident in Fig.4.12 and Fig.4.13 for all the schemes H.263+ inter-frame coding, STD 2D-TDWZ, and proposed 3D-TDWZ. For 3D-TDWZ, since even less number of bits used in WZ coding than that in 2D-TDWZ due to the additional exploitation in temporal direction, the compression performance is enhanced significantly comparing to 2D-TDWZ. For both tested sequences, there is around 1dB difference in PSNR between the 3D-TDWZ with GOP length $n=4$ and $n=8$. The 3D-TDWZ with GOP length $n=4$ even outperforms the 2D-TDWZ with GOP length $n=8$.

The enhancement due to the utilisation of 3D DCT in TDWZ is significant in above simulations. Comparing to the 2D-TDWZ and H.263+ Intra-frame coding, the computation complexity of the proposed 3D-TDWZ is slightly higher than that of the 2D-TDWZ due to the operation of the third 1D DCT transformation in temporal direction, and similar as that of H.263+ DCT intra-frame coding. The overall complexity is low which means 3D-TDWZ is suitable to be applied in the applications with low complexity encoder. Next section will reveal the performance of transform domain distributed multiview video coding which applies 2D-TDWZ and 3D-TDWZ.

4.4 Transform Domain Distributed Multiview Video Coding

4.4.1 Review of DMVC

One of the most important applications of WZ video coding is multiview video coding. A multiview system refers to a system where multiple cameras are monitoring the same scene from different viewing positions. The applications of multiview systems are very wide and cover different areas such as homeland security, military, environment monitoring and healthcare.

The multiview video coding algorithm concerns how to effectively compress the video content from multiple cameras. Basically, there are three ways to compress multiple video sources:

Case 1: each camera is encoded separately by the conventional video coding standard and separately decoded. The architecture is illustrated as Fig.4.14.

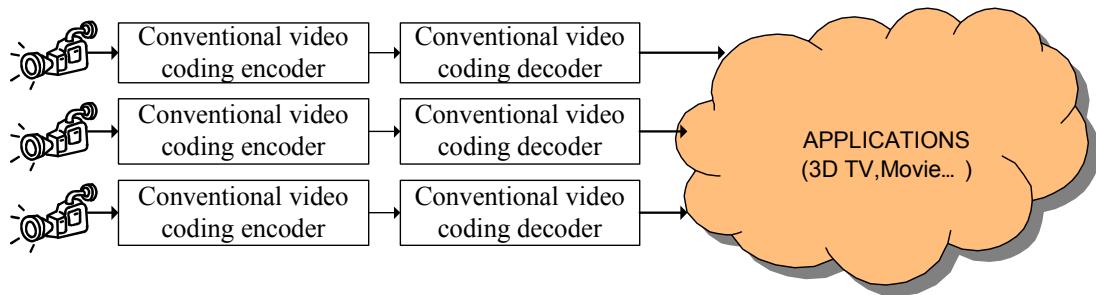


Fig.4. 14 Multiview video coding case 1(conventional coding)

Case 2: in this case, the video content from multiple cameras are jointly encoded and jointly decoded. The typical configuration is one camera is encoded via conventional predictive video coding, the other cameras perform ME and motion compensation with this camera and only the residual and MV are encoded. This is called interview coding. Fig. 4.15 and Fig. 4.16 illustrate the architecture and a classic interview coding structure introduced in [118].

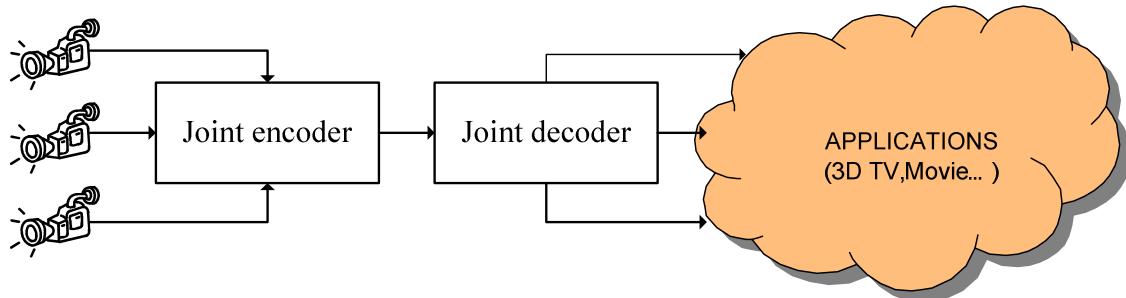


Fig.4. 15 Multiview video coding case 2(Interview coding)

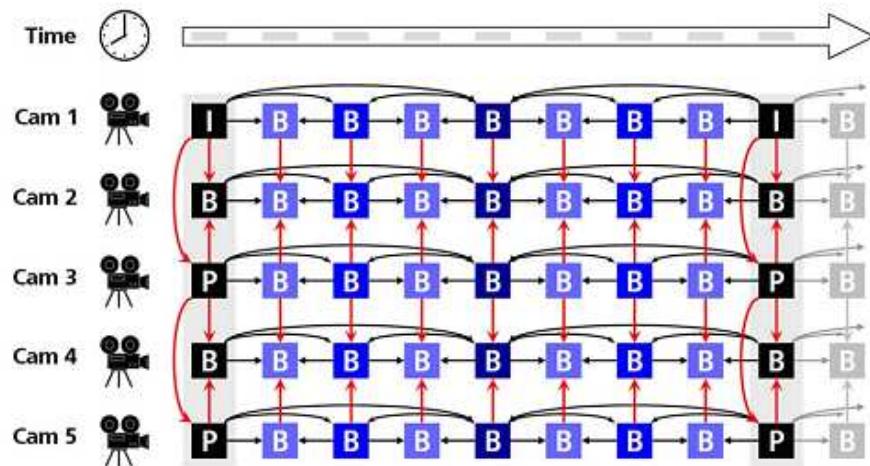
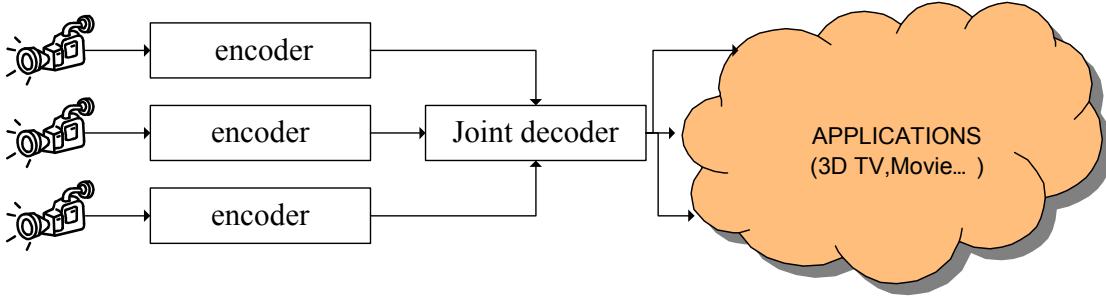


Fig.4. 16 A typical interview video coding structure [118]

Case 3: this case is referred as Distributed Multiview Video Coding (DMVC), in which each camera is separately encoded but jointly decoded. DMVC proposed in [100, 103, 119-121] employs the WZ coding to compress multiview video. A typical architecture of it is shown in

Fig.4.17.

**Fig.4. 17 Multiview video coding case 3(Distributed multiview video coding)**

Since in multiview video coding, these cameras capture the same scene from different angles, high correlation exists among these cameras. For Case 1, each encoder employs the conventional video coding standard and works separately, the correlation among cameras is not exploited at all thus the compress is not efficient. For Case 2, the correlation is exploited by the interview prediction thus the best compression can be provided. However, complex encoders are required in both cases due to the predictive coding structure. It is not realistic in certain applications whose encoders are energy constrained and cannot afford huge task. Particularly, in Case 2, the prediction structure requires the cameras to be able to communicate with each other, which is not available in multiview coding network, where cameras are deployed at different positions and unable to communicate with other cameras. Moreover, communication among cameras causes data exchange which is also not feasible in practice. Case 3 provides an efficient multiview video coding architecture: Firstly, it doesn't need a complex encoder to perform encoding. Secondly, it doesn't require cameras to communicate with each other during the encoding. Thirdly, it exploits the correlation among the cameras at the decoder thus the compress efficiency still can be achieved. The DMVC is suitable for certain applications with a low complexity encoder.

Many previous literatures regarding to DMVC are mainly based on pixel domain WZ video coding principle apart from DMVC introduced in [100, 101] where 2D DWT is considered. In the thesis, we focus on investigating the performance of DCT transform domain DMVC. Specifically, three types of transform domain DMVC based on 2D-TDWZ or 3D-TDWZ are proposed. The later sections are organised as following: in next section, the basic setup including camera configuration and joint side information generation for DMVC are introduced. Then the pixel domain DMVC, transform domain DMVC based on 2D-TDWZ, transform domain DMVC based on 3D-TDWZ and transform domain residual DMVC based on 3D-TDWZ are presented subsequently including architectures and operations. The simulation results of all these introduced DMVC issues are analyzed and compared

4.4.2 Configuration in DMVC

Before presenting the proposed DMVC, the basic setup for DMVC including camera configuration and joint side information generation is presented in this section. Camera configuration determines the property of the camera and the coding way of the video frames inside. The joint side information generation process is used to generate the corresponding side information for the frame coded by WZ coding.

4.4.2.1 Camera Configuration

DMVC based on WZ video coding is a relatively new concept proposed by Guo in his work [100] in 2006. Later in [36, 103, 121, 122], different DMVC architectures are presented. There are typically three types of camera configuration in above works and the below

I) Type I presented in [100]

In this type of camera configuration, all cameras have same prosperities and encode frame by intra-frame coding (I frame) or WZ coding in turn as shown in Fig.4.18. The GOP length $n=2$.

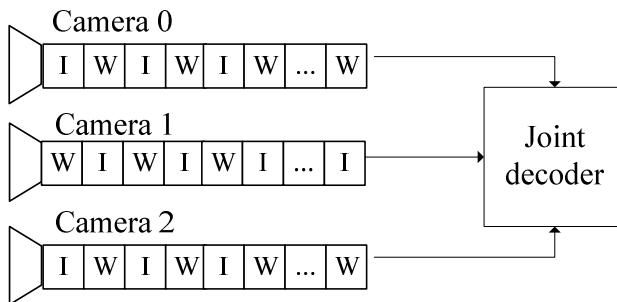


Fig.4. 18 Frame composition of Key camera and WZ camera in [100]

II) Type 2 presented in [36]

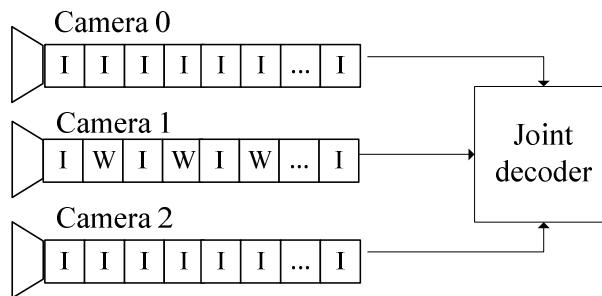


Fig.4. 19 Frame composition of Key camera and WZ camera in [36]

In this type of camera configuration, there are two types of cameras: intra camera and WZ camera. In Fig.4.19, the camera 0 and camera 2 are intra cameras, in which all the frames are coded in conventional way. Camera 1 is WZ camera, in which half number of frames are I frames which are coded by intra-frame coding and half number of frames are WZ frames which are coded by WZ coding.

III) Type 3, presented in [103]

In this case, cameras are divided into two types: key camera and WZ camera. All frames in key camera are coded by intra-frame coding and in WZ camera, some frames are coded by intra-coding and the rest are WZ frames which are coded by WZ coding. In Fig.4.20, camera 1 is key camera and camera 0 and 2 are WZ cameras.

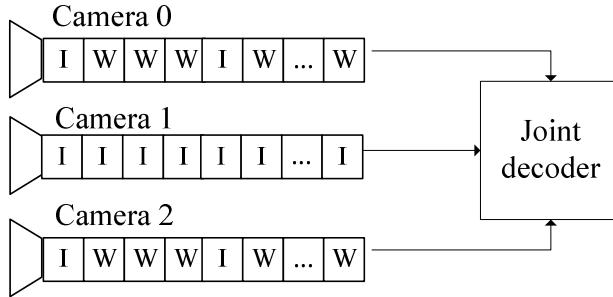


Fig.4. 20 Frame composition of Key camera and WZ camera in [103]

In this thesis, camera configuration of type III) is used to build up the DMVC. Comparing these three types of camera configuration, type III) offers the best compression performance followed by type I) and type II). Type III) uses larger GOP and has less number of frames to be coded by the intra-frame coding than type I) and II). However, type II) offers the best side information for each WZ frame since more I frames are involved to generate the side information than type I) and III) but brings down the overall compression effect since I frames are coded by intra-frame coding.

4.4.2.2 Joint Side Information Generation

With the camera configuration type chosen, this part describes the corresponding side information generation for the WZ frame in WZ camera. In WZ video coding, side information is interpolated by using key frames from the same camera. This is called intra-frame interpolation. In DMVC, frames from neighbouring camera can be used to assist the side information generation. The side information generation using only frames from neighbouring camera is called inter-camera interpolation (or interview interpolation). It is proved in [36, 100, 103, 104] that using the combination of intra camera interpolation and inter camera interpolation can provide better side information than only using any single interpolation scheme. Intra camera Interpolation is not good at estimating high motion areas and inter camera Interpolation has problem with scene occlusions and reflections. Thus, in this chapter, we propose a joint side information generation for the camera configuration type chosen. The general structure of joint side information is shown in Fig.4.21. As introduced in the part of camera configuration, $W_1W_2W_3W_4$ are frames from WZ camera, in which W_1 is intra-frame coded and the rest are WZ frames to be coded by the WZ

coding. $K_1K_2K_3K_4$ are key frames from key camera which are all coded by intra-frame coding. Y_2 is the corresponding side information of W_2 which is jointly generated by using $W_1K_1K_2$. Y_3 is generated by using $W_1K_1K_3$ and Y_4 is obtained by using $W_1K_1K_4$.

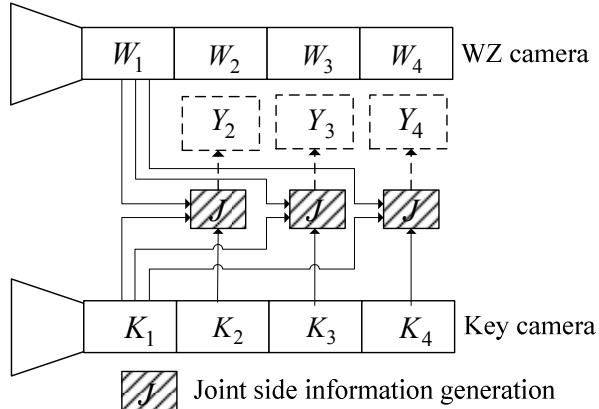


Fig.4. 21 The architecture of joint side information generation

Fig.4.22 shows the exact procedure of f generating side information Y_2 by joint interpolation and the details are given as following.

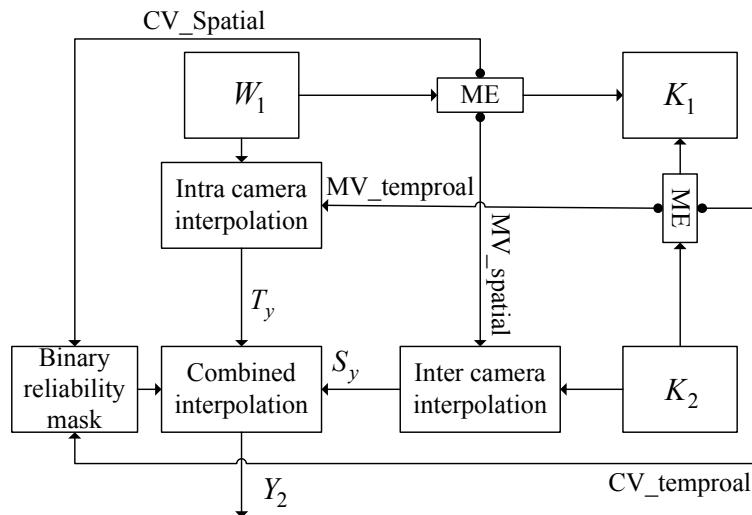


Fig.4. 22 Side information generation in JSG

- Firstly, the temporal motion vectors indicated by $MV_{temporal}$ together with a cost vectors indicated by $CV_{temporal}$ are calculated between the frames K_1 and K_2 from key camera based on block matching. Here the cost list contains the difference strengths of each pair blocks associated in the motion vectors and the Mean Square Difference (MSD) value is used to describe the cost of each pair blocks.
- Another group of motion vectors indicated by $MV_{spatial}$ together with a cost vectors indicated by $CV_{spatial}$ are calculated subsequently between the frame K_1 from key camera

and the intra frame W_1 from WZ camera.

- Based on these two cost vectors, a binary reliability mask is designed. For each position, if the value of CV_temporal is bigger than CV_spatial we set 1 otherwise we set 0.
- The first estimated version of side information T_y is obtained by utilizing MV_temporal and the frame W_1 .
- We also get the second estimation of side information S_y by utilizing MV_spatial and K_2 .
- Finally, according to the mask vectors, Y_2 is generated by choosing values from T_y or S_y based on the binary reliability mask.

Fig.4.23 illustrates the side information comparison generated via three interpolation algorithms for Vassar sequence. Note the difference in the area marked by white circle. Neither single intra-camera interpolation nor single inter-camera interpolation can give better output than using joint interpolation.

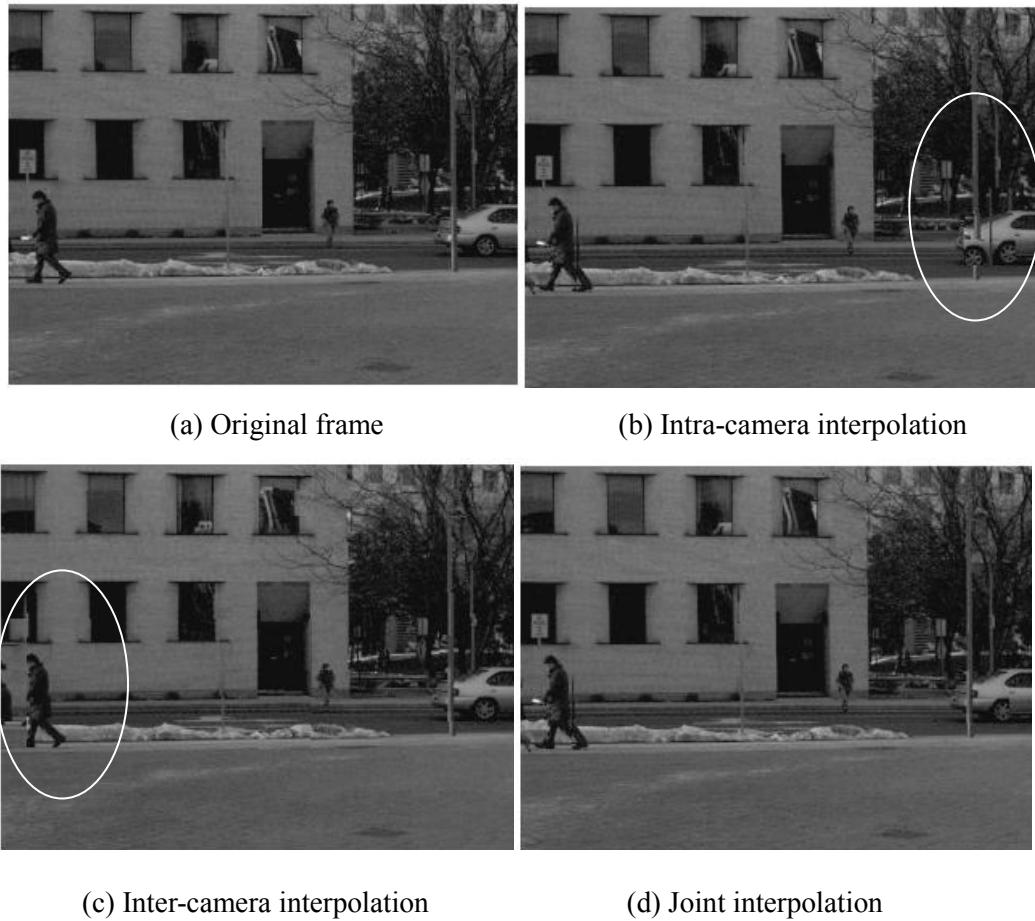


Fig.4. 23 Side information comparison for different interpolation in multiveiw

4.4.3 Pixel Domain DMVC

In this section, a generic pixel domain DMVC (PD-DMVC) is presented as introduced in [103],

based on which the transform domain DMVC proposed in this thesis is built in later sections. Note the architecture may not be fully same as proposed in [103] but basic camera configuration and coding process keeps similar. Fig.4.24 shows the architecture of the pixel domain DMVC including one key camera and one WZ camera.

The encoding process is explained as following steps.

- In key camera, all frames are coded via conventional intra-frame coding. In WZ camera, the frames are organised into GOPs with size of n . We denote the frames from WZ camera as $W_{g,t}$, where g represents the index of GOP and $t(1 \leq t \leq n)$ represents the temporal index of frames inside the GOP. $K_{g,t}$ denotes the corresponding key frame in key camera with same temporal instant as $W_{g,t}$. In each WZ camera, $W_{g,1}$, the first frame of each GOP, is I frame and intra coded and the rest frames $W_{g,2} \dots W_{g,n}$ are WZ frames and coded frame by frame via WZ coding.
- For a WZ frame (take $W_{g,2}$ as example), an uniform quantisation is applied with 2^M levels where $M=\{1,2,3,4\}$. After the quantisation, the quantised symbols Q_k will be converted into M bits binary codewords.

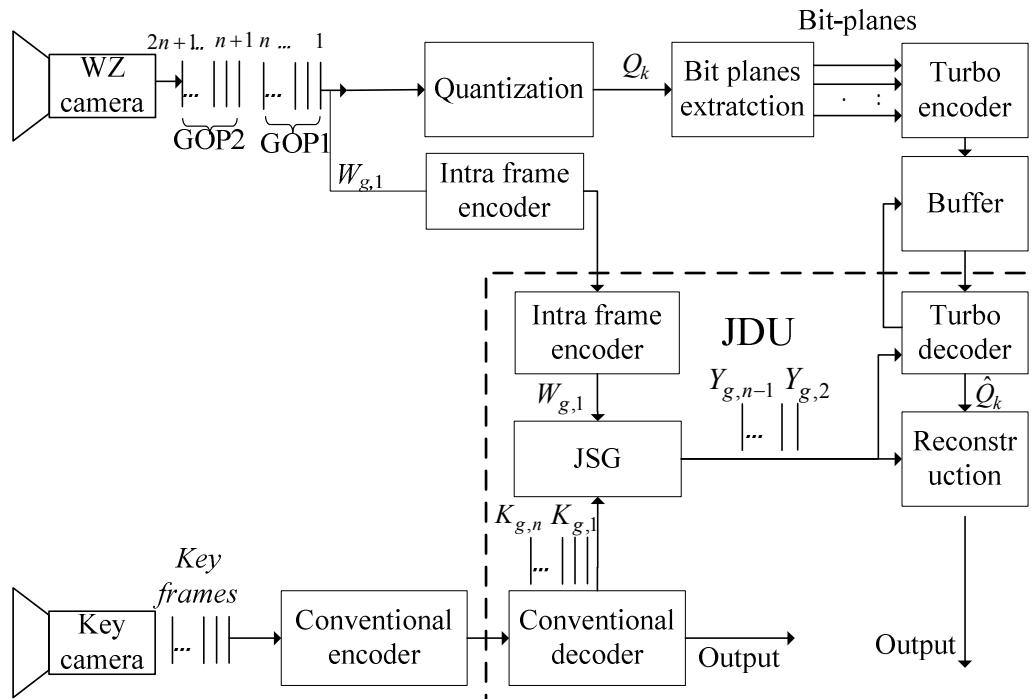


Fig.4. 24 The architecture of pixel domain DMVC

- The bit-plane extraction is performed to form M bit-planes according to the significance of the bit in binary codeword. The bits with same significance are blocked together to form a bit-plane.
- The most significant bit-plane will be first sent to Slepian-wolf codec for encoding. The

next bit-plane will start to transmit only after previous bit-plane is successfully received at the decoder. The Slepian-Wolf codec is implemented by a rate-compatible punctured turbo code (RCPT) in combination with a feedback channel.

- After the turbo encoding, the generated parity bits are saved in a buffer while the systematic bits are discarded. A small subset of these parity bits is transmitted to the decoder for decoding.

The decoding is done in a joint decoding unit (JDU) and the decoding process of $W_{g,2}$ involves following steps.

- The key frame $K_{g,1}$ and $K_{g,2}$ from key camera and $W_{g,1}$ from WZ camera are first decoded in the conventional way.
- $K_{g,1}$, $K_{g,2}$ and $W_{g,1}$ are inputted to the unit of Joint Side information Generation(JSG) to mutually generate the corresponding side information $Y_{g,2}$ for $W_{g,2}$. The correlation between $Y_{g,2}$ and $W_{g,2}$ is modelled by Laplacian variable. The corresponding Laplacian parameter can be calculated offline or by using $K_{g,1}$ instead.
- With $Y_{g,2}$, received parity bits and Laplacian parameter , the turbo decoding is performed and the transmitted bit-plane can be decoded. If the BER of the decoded bit-plane is bigger than a predefined threshold, e.g. 10^{-3} , a request for more parity bits will be sent back to the encoder. The encoder will send more parity bits until the whole decoding is successfully done.
- After all the bit-planes are decoded, the quantised symbols \hat{Q}_k are reconstructed.
- Frame reconstruction is performed according to $E(\hat{W}_{g,2} | \hat{Q}_k, Y_{g,2})$ and next frame $W_{g,3}$ will start to transmit.

4.4.4 Transform Domain DMVC with 2D DCT

Fig. 4.25 illustrates the architecture of transform domain DMVC based on 2D DCT (2TD-DMVC). This architecture is developed based on PD-DMVC by apply 2D DCT TDWZ. In the architecture, the configuration for key camera and WZ camera keep same definition as in PD-DMVC. However, the encoding process is different, which is explained below.

- Key frames and $W_{g,1}$ are encoded by intra-frame coding.
- For each WZ frame (for example, $W_{g,2}$), a 4x4 blockwise DCT transformation is applied to remove spatial redundancy. Coefficients at same position in DCT block are grouped into one coefficient band. There are total 16 Coefficient bands are obtained.
- For a coefficient band $C_{w,k}$, where w indicates the coefficient band for WZ frame and k is

coefficient band index with $1 \leq k \leq 16$, the uniform quantisation is applied. The quantisation level is determined by predefined quantisation matrix. After the quantisation, the quantised symbol Q_k is converted to binary codewords.

- Bit-plane extraction is performed and bit-planes are turbo encoded and only part of the parity bits is sent.
- Inside the JDU, $Y_{g,2}$ is firstly generated by using $W_{g,1}$, $K_{g,1}$ and $K_{g,2}$ in JSG.
- 4x4 blockwise DCT is applied to the $Y_{g,2}$ and coefficient bands are grouped similarly as done at the encoder. Corresponding coefficient band $C_{y,k}$ of side information is used to decode the current bit-plane together with the received parity bits.
- Same as done in PD-DMVC, the decoding process repeats until the current bit-plane is decoded successfully. After all the bit-planes decoded, the quantised symbol \hat{Q}_k is reconstructed and the coefficient band will be reconstructed as $E(\hat{C}_{w,k} | \hat{Q}_k, C_{y,k})$
- The inverse DCT transformation is applied after all coefficient bands are received and the frame is reconstructed. $W_{g,3}$ will start to transmit.

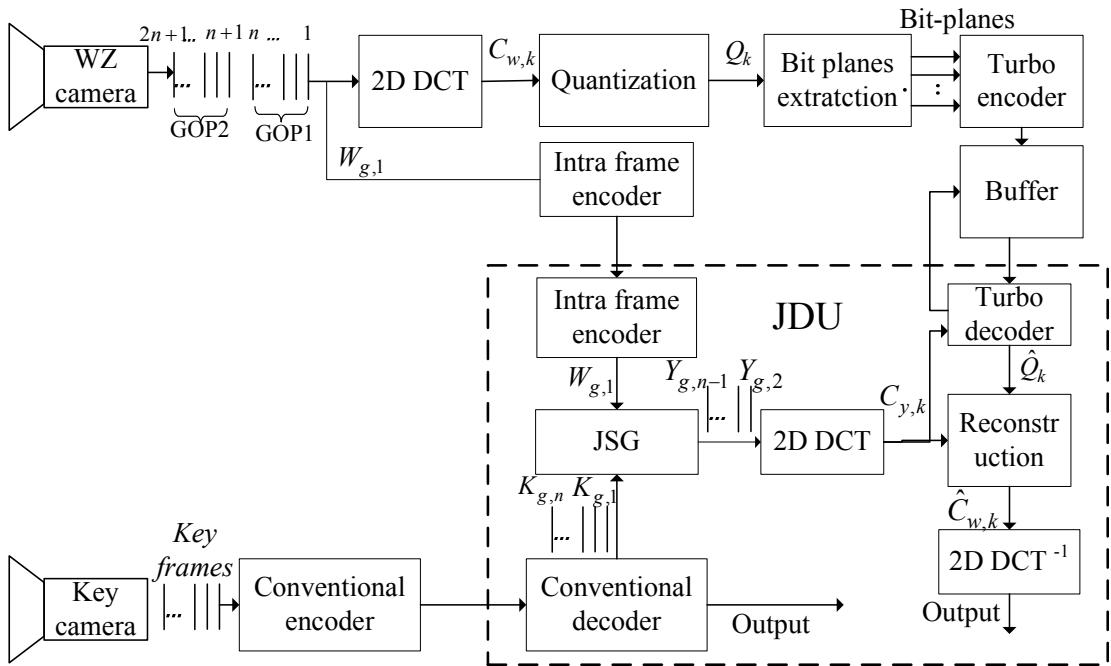


Fig.4. 25 The architecture of transform domain DMVC with 2D DCT

The quantisation is operated based on the quantisation matrix. We use the quantisation matrices of [31] to get different R-D points in simulation. The used quantisation matrices are shown in Fig.4.26, in which each element in the quantisation matrix determines the quantisation levels for the coefficient band located at same position in DCT matrix.

\bar{M}^1	\bar{M}^2	\bar{M}^3	\bar{M}^4	\bar{M}^5	\bar{M}^6	\bar{M}^7
64 32 16 8 32 16 8 4 16 8 4 4 8 4 4 0	64 16 8 8 16 8 8 4 8 8 4 4 8 4 4 0	32 16 8 4 16 8 4 4 8 4 4 0 4 4 0 0	32 16 8 4 16 8 4 0 8 4 0 0 4 0 0 0	32 8 4 0 8 4 0 0 4 0 0 0 0 0 0 0	32 8 0 0 8 0 0 0 0 0 0 0 0 0 0 0	16 8 0 0 8 0 0 0 0 0 0 0 0 0 0 0

Fig.4. 26 Quantisation matrix for transform domain DMVC with 2D DCT [31]

4.4.5 Transform Domain DMVC with 3D DCT

Based on 2TD-DMVC, a transform domain DMVC with 3D DCT (3TD-DMVC) is proposed and the architecture is illustrated in Fig.4.27. This work is the extension of the 2TD-DMVC, in which 3D-TDWZ introduced in the Section 4.3 are applied. Same as above DMVC issues, we take one key frame camera and one WZ camera as an example to illustrate the coding process. In WZ camera, the frames are organised into GOPs with size of n .

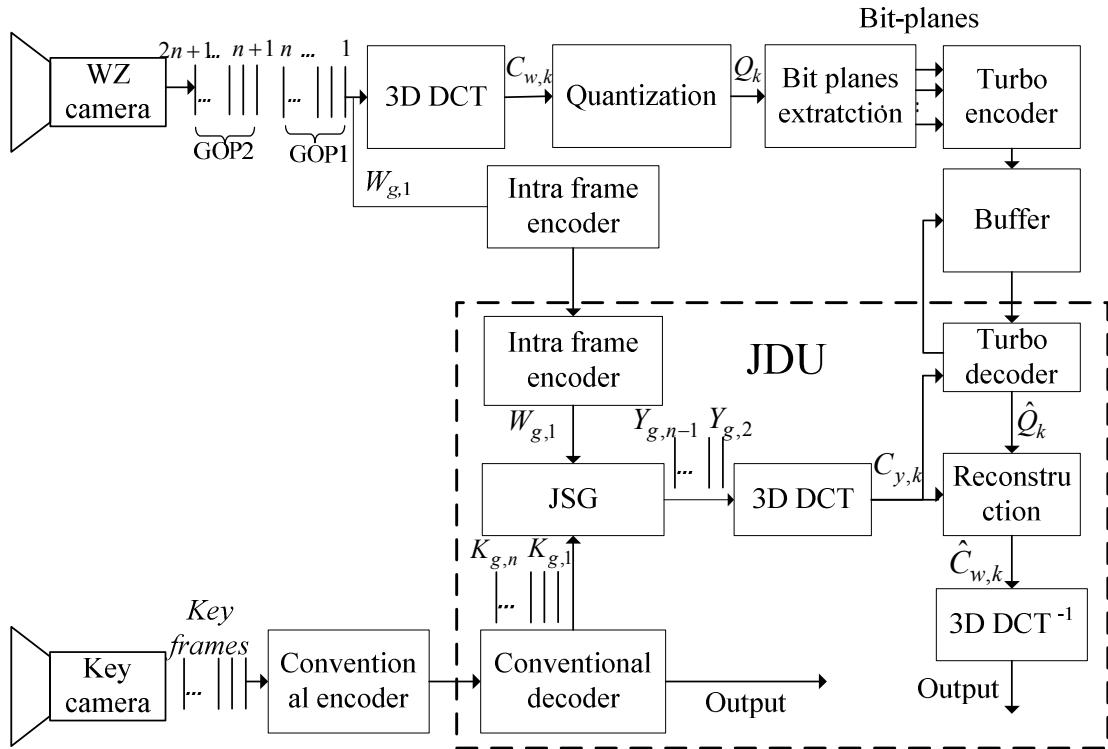


Fig.4. 27 The architecture of transform domain DMVC with 3D DCT

- $W_{g,t}$, $K_{g,t}$, n keeps same definition as in Section 4.4.3 and Section 4.4.4.
- In each WZ camera, $W_{g,1}$, the first frame of each GOP, is I frame and intra coded and the rest frames $W_{g,2} \dots W_{g,n}$ are WZ frames and coded in the way of transform domain WZ coding with 3D DCT. These WZ frames $W_{g,2} \dots W_{g,n}$ are grouped together then a $(m, m, n-1)$

cubewise 3D DCT is applied.

- After the transformation, the Coefficients with same positions in every DCT cube are grouped together to compose the coefficient bands, denoted by $C_{w,k}$ (w indicates the coefficient band for WZ frame and k is coefficient band index with $1 \leq k \leq m \times m \times (n-1)$).
- Each coefficient band $C_{w,k}$ is quantised using a uniform scalar quantiser with 2^{M_k} levels and the quantisation levels for each coefficient band is determined by the predefined 3D quantisation volume (3D quantisation matrix). The design of quantisation volume is discussed in Section 4.3.4.
- After the quantisation, the quantised symbols Q_k are converted into fixed-length binary code words and corresponding bit-planes extraction is performed. Each bit-plane is then sent to the Slepian-Wolf encoder for encoding.
- After encoding, the parity bits produced by the turbo encoder are stored in a buffer then a subset of these parity bits is transmitted to the decoder for decoding.
- Inside the JDU at the decoder, $k_{g,1} \dots k_{g,n}$ from key camera are firstly decoded. Meanwhile, $W_{g,1}$ is also intra decoded. With $k_{g,1}, k_{g,2}, k_{g,3} \dots k_{g,n}$ and $W_{g,1}$, the corresponding side information $Y_{g,2}, Y_{g,3}, \dots Y_{g,n}$ are generated in the JSG.
- With generated side information group, $m \times m \times (n-1)$ cubewise 3D DCT is applied.
- Same as done in WZ camera, the coefficient band $C_{y,k}$ of the side information is grouped and used to help turbo decoder to decode the bit-plane with received parity bits. The Laplacian distribution is used to model the correlation between $C_{y,k}$ and $C_{w,k}$.
- If decoding is not successful, namely the BER is larger than a predefined threshold, e.g. 10^{-3} , the request for more parity bits will be sent back to the encoder via feedback channel. The encoder sends more parity bits based on the request.
- After all bit-planes are decoded, quantised symbol \hat{Q}_k can be obtained. With \hat{Q}_k , the reconstruction of the coefficient band $\hat{C}_{w,k}$ is performed, which follows the same model as above by $E(\hat{C}_{w,k} | \hat{Q}_k, C_{y,k})$.
- After all coefficient bands are reconstructed, inverse 3D DCT is performed and $W_{g,2} \dots W_{g,n}$ is decoded and then next GOP will start to transmit.

4.4.6 Transform Domain Residual DMVC with 3D DCT

In [32], Arron explored the performance of WZ coding of residual frames in pixel domain. In that work, WZ frame is not encoded instead the residual frame between WZ frame and the reference frame is encoded. The residual frame is obtained by making simple subtraction in order to keep low complexity of the encoder. The reference frame is available at both encoder and decoder. At

the decoder, the WZ frame is reconstructed by using decoded residual frame and the reference frame. The performance of PDWZ is much enhanced and even can reach similar level as that of TDWZ with the help of hash code generated at the encoder. In [123], the case of WZ residual coding in transform domain is considered, in which superior performance is reported to the normal TDWZ. Both works assume that the reference frame used to produce the residual frame is available at both the encoder and decoder. The principle of residual frame work is similar as DPMC. Because of the high correlation in temporal direction, the residual value tends to be zero or very smaller value than original pixels. When same number of quantisation levels applied, smaller distortion is obtained after the reconstruction.

In DMVC introduced in Section 4.4.3 and Section 4.4.4, we can just use the first frame of each GOP in WZ camera which is intra-frame coded as reference frame to produce the residual. The idea of residual coding can be exactly applied in DMVC and extended by using 3D DCT. In this section, we propose a transform domain residual DMVC using 3D DCT (3TDS-DMVC). The proposed architecture is shown in Fig.4.28.

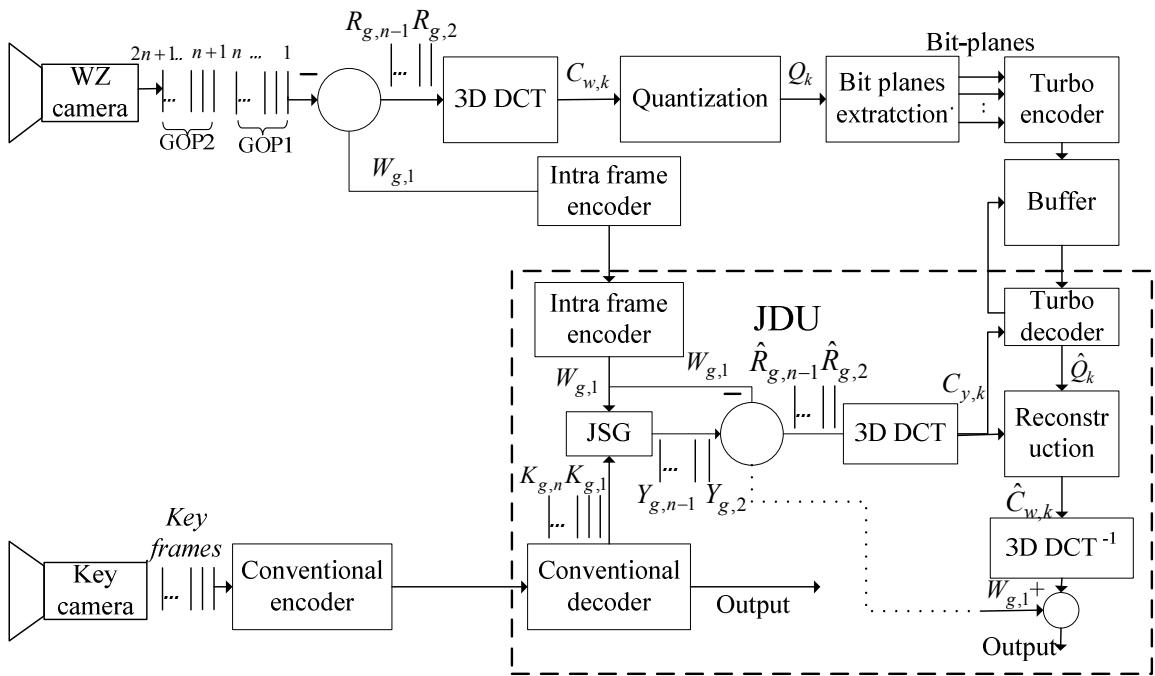


Fig.4. 28 The architecture of Transform domain residual DMVC with 3D DCT

This architecture can be considered as the extension of that of 3TD-DMVC. In this system, key camera operates same as that of 3TD-DMVC and WZ camera still groups the frames into GOPs for coding. The coding process is mostly similar as done in 3TD-DMVC. However, there are some different operations which are mentioned below.

- All the variables keep same definition as in 3TD-DMVC.

- Unlike as done in 3TD-DMVC, we do not apply 3D DCT to the rest WZ frames $W_{g,2} \dots W_{g,n}$, but apply 3D DCT to the residual frames $R_{g,2} \dots R_{g,n}$, which are obtained by making subtraction between $W_{g,2} \dots W_{g,n}$ with $W_{g,1}$ respectively.
- After 3D DCT transformation, the coefficient bands are grouped and scalar quantisation is applied. The quantisation level for each coefficient band is exclusively determined by the quantisation volume designed in later section.
- Then bit-plane extraction, Slepian-Wolf encoding and parity bits transmission are operated subsequently.
- At the decoder, with received key frames and $W_{g,1}$, the side information frames $Y_{g,2}, Y_{g,3} \dots Y_{g,n}$ are generated inside the JSG.
- As done at the encoder, the residual side information frames are obtained by making subtraction between $Y_{g,t}$ and $W_{g,1}$, which then are used to help turbo decoder to decode bit-planes with received parity bits.
- Once all coefficient bands are decoded, they are grouped together and inverse 3D DCT is applied to get the decoded WZ residual frame group.
- Adding WZ residual frames with $W_{g,1}$, the WZ frames are reconstructed .

4.4.7 Results and Discussions

In Section 4.4.3, Section 4.4.4, Section 4.4.5 and Section 4.4.6, we gradually introduced 4 types of DMVC issues from the simplest PD-DMVC to 2TD-DMVC, 3TD-DMVC and 3TDS-DMVC. In the simulation, the performance of all these four architectures are presented together to make comparison. In addition, H.263+ intra-frame coding is also added as reference. Two multiview sequences “Ballroom” and “Vassar” from Mitsubishi Electric Research Laboratories (MERL) with resolution level of 320x240 and frame rate of 30fps are used.

- For both sequences, camera 1 is used as key camera and camera 0 is used as WZ camera. The key frames from key camera and intra frame from WZ camera are assumed to be reconstructed perfectly in JDU.
- For each camera, 48 frames are investigated. The GOP size is set to 4 thus there are total 36 WZ frames from WZ camera compressed. Simulation results reveal the average PSNR performance of Y component of these 36 WZ frames.
- The block size is 4 for 2D DCT transform while the cube size is 4x4x3 for 3D DCT.
- The quantisation volume for 3D DCT is designed with parameters $\beta_i=0.001$, $\beta_o=0.3$ and C varies in [1 2 3 4 5 6 7] thus to get different RD points. For 2TD-DMVC, 7 quantisation matrices from [31] are used. For PD-DMVC, $2^M=[2\ 4\ 8\ 16]$ quantisation levels are used.

- For all architectures, turbo codec is composed by two identical constituent convolutional encoder of rate 1/2 with constraint length of 4 and polynomial generator of $g = (13,11)$. The puncturing period is set to 8 thus to provide the coding rate of $\{0,8/9,8/10,8/11\dots1/3\}$.

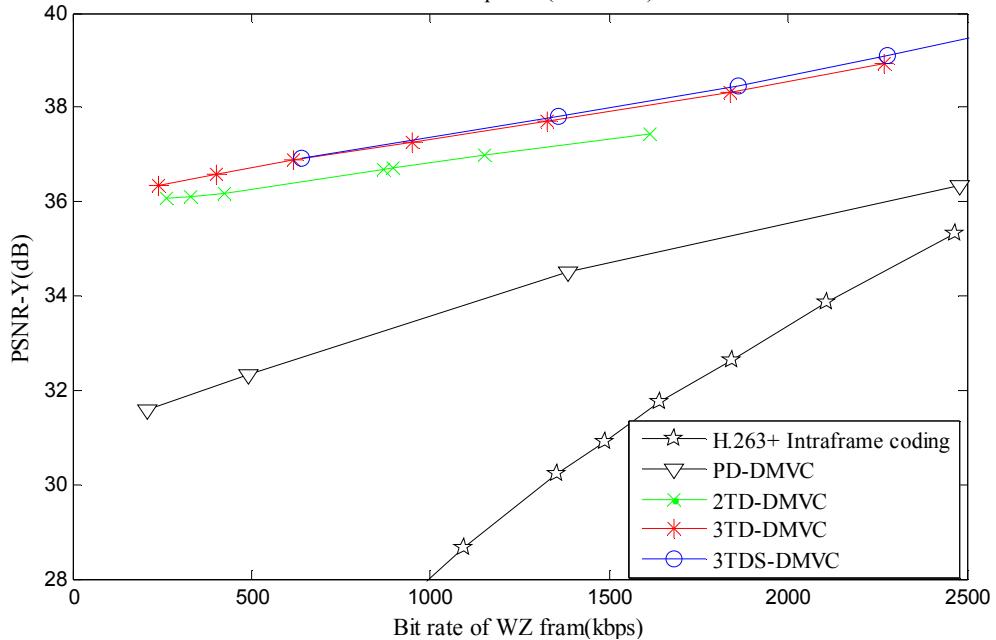


Fig.4.29 R-D performance of proposed transform domain DMVC for Vassar sequence (Camera 0)

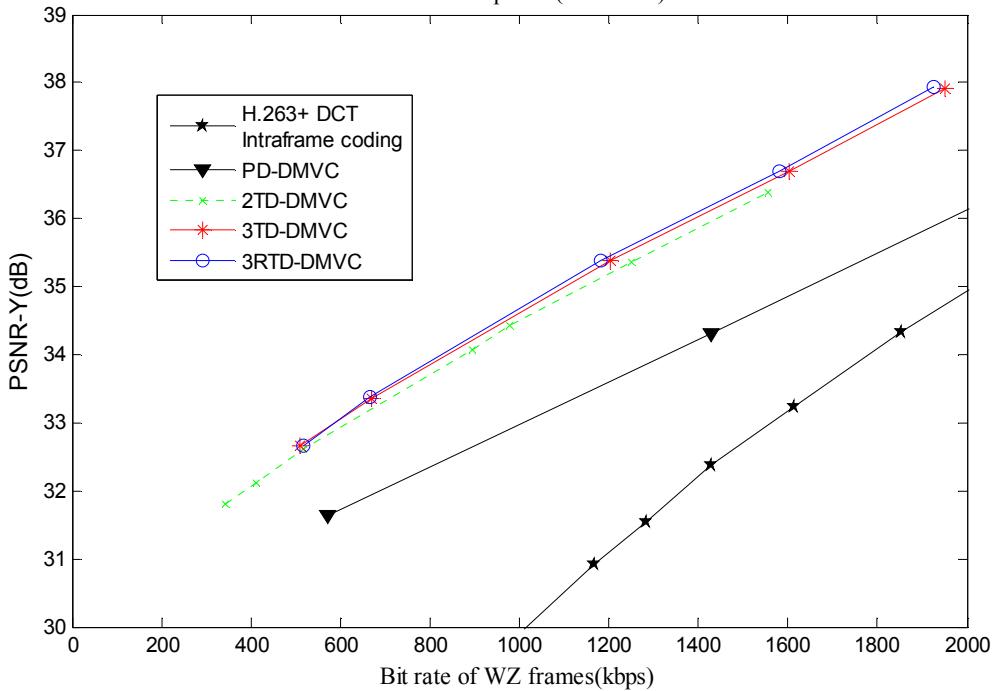


Fig.4.30 R-D performance of proposed transform domain DMVC for Ballroom sequence (Camera 1)

In Fig.4.29 and Fig.4.30, it is obvious that all introduced DMVC issues significantly outperforms the H.263+ DCT based intra-frame coding. For the applications with restricted computation of the encoder, only multiview video coding Case 1 (separate encoding and separate decoding) and Case 3(DMVC) introduced in Section 4.4.1 are feasible. For Case 1, all cameras have to perform intra-frame coding e.g. H.263+ intra-frame coding for the purpose of low complexity. Fig.4.29 and Fig.4.30 show the advantage of DMVC over Case 1. DMVC achieves higher compression performance than Case 1 using intra-frame coding whilst DMVC keeps similar complexity as Case 1. Moreover, the proposed transform domain DMVC further improves the system performance significantly comparing to PD-DMVC. A clear gap can be observed between PD-DMVC and 2TD-DMVC, which indicates the maximum of 1.2 dB differences of PSNR for “Ballroom” sequence and 4.4 dB for “Vassar” sequence.

Within the transform domain DMVC solutions, 3TD-DMVC outperforms the 2TD-DMVC duo to the exploitation in the temporal direction among WZ frames. For “Vassar” sequence, 3TD-DMVC improves PSNR gain up to 0.6dB. In the case of “Ballroom”, the PSNR is enhanced around 0.3dB maximum. “Ballroom” sequence has lower correlation in temporal direction than “Vassar” sequence due to the fast motion inside thus 3D DCT based algorithms improve PSNR of “Ballroom” sequence less significantly than that of “Vassar” sequence.

It also can be observed that 3TDS-DMVC can slightly further enhance the system performance than 3TD-DMVC. The system enhancement is around 0.2dB for “Vassar” sequence and 0.02dB for “Ballroom” sequence. For the sequence with fast motion (low temporal correlation), the 3TDS-DMVC performs similarly as 3TD-DMVC. For this type of sequence, the residual value between pixels tends to be very large thus the residual coding technique cannot show obvious advantage in this case.

The complexity of the encoder is gradually increased with above introduced DMVC solutions. The simplest case is PD-DMVC, followed by the 2TD-DMVC. The encoder complexity of 3TD-DMVC is slightly higher than that of previous two solutions and is similar as that of conventional intra-frame coding algorithm. 3TDS-DMVC increases little complexity of the encoder comparing to the 3TD-DMVC due to the subtraction made to obtain the residual. However, its encoder is still within the range of low complexity.

4.5 Chapter Summary

In this chapter, we first investigate the architecture of TDWZ with 3D DCT. Due to the lack of exploitation in temporal correlation in current WZ video coding, 3D DCT is introduced to perform such task as what ME does in traditional predictive coding. In the proposed architecture,

WZ frames are grouped into GOW and each GOW is transformed by 3D DCT. In the simulation, it is proved that by using 3D DCT, the system performance is considerably enhanced comparing to that of the TDWZ based on 2D DCT.

Moreover, we investigate the performances of transform domain DMVC, which is one of the most applications of WZ video coding paradigm. DMVC is suitable to be applied in the multiview video applications with restricted encoder in computation and complexity. Most previous available issues regarding to DMVC are based on pixel domain. Due to the exploitation in spatial and temporal direction, the proposed transform domain DMVC can further improve the system performance significantly. In this chapter, we mainly investigate three types of transform domain DMVC, which are 2D DCT based DMVC, 3D DCT based DMVC and 3D DCT based residual DMVC. The simulation results clearly show the significant improvement of transform domain DVC comparing to the normal pixel domain DMVC and DMVC using conventional intra-frame coding. Besides, it also shows 3D DCT based DMVC outperforms the 2D DCT based DMVC in general. By applying 3D DCT to the residual frames in DMVC, the compression performance can be further enhanced slightly with no more than 0.2dB gain in PSNR for the test sequences comparing to the DMVC based on 3D DCT.

After the research work of WZ video coding in pixel domain, transform domain and its application in DMVC, the next chapter will address another important application of WZ coding which is error-resilience of wavelet video transmission over error-prone channel.

Chapter 5

5. Efficient Error Resilient Scheme for Wavelet Video Codec using Wyner-Ziv Coding

One of the important applications for WZ coding is robust transmission for video coding. Previous literatures have explored the applications of WZ coding in H.264 or MPEG video transmission. This chapter will bring a new WZ coding scheme called WZER especially for wavelet based video codec. The proposed scheme emphasises on the protection of the Region Of Interest (ROI) area in the frame thus to get the trade-off between error resilience and bandwidth usage. The performance of proposed WZER is investigated under two different error prone channel: AWGN channel and packet erasure channel. The simulation results showed the advantages of WZER in the aspects of bandwidth usage and ROI region recovery comparing to the traditional FEC scheme.

5.1 Review of Error-Resilient Techniques for Wavelet Video Codec

With the rapid growth of internet and modern communications technology, the strong demand of various multimedia applications and services has promoted the research of robust transmission of compressed image and video data. In recent years, the design of robust video transmission techniques over heterogeneous and unreliable channels has been an active research area. However, error control in image/video communications is proved to be a very tough task. In most popular video coding standards, compressed video streams are very susceptible to transmission errors due to the utilisation of spatial-temporal prediction coding structure and Variable Length Coding (VLC) at the source encoder. A single incorrectly recovered pixel in reference frame can lead to errors in the samples of the following frames which are inter-frame coded. Therefore, in order to deliver the compressed bit stream over an error-prone network, the video has to be coded in a resilient format to combat the channel errors. In [124], Wang et al reviewed most error-resilient techniques and categorized them into three groups: Firstly, working at source and channel coder to make the bit stream more error-resilient; Secondly, promoting the error concealment method at the decoder such as using interpolation techniques etc.; Thirdly, introducing the interaction between the encoder and decoder so that encoder can adapt the operation based on loss information provided by decoder. A feedback channel working with ARQ is usually considered for this algorithm.

In most video coding standards, the Discrete Cosine Transform (DCT) is widely used to realise

spatial compression. However, the later emerging Discrete Wavelet Transform (DWT) has become increasingly popular in last decades since Shapiro [47] and Said [48] introduced the Embedded Zerotree Wavelet (EZW) and Set Partitioning In Hierarchical Trees(SPIHT) to efficiently code the wavelet coefficients. Later in [125] and [126], 3D-EZW and 3D-SPIHT were proposed to provide more superior performance. DWT has distinct advantages in reducing the block artifacts especially at the low bit rate and inherited scalability characteristic comparing to the DCT transformation. With these advantages of DWT, the wavelet based video coding combining DWT transformation and block based predictive coding has been proposed in various applications. However, error-resilient ability of video stream has to be considered as well in order to extend the application area of such wavelet based video coding. Comparing to the DCT based video coding algorithms, DWT based video coding will experience more problems due to the use of EZW or SPIHT. Not only it will suffer same problems resulted by predictive coding and VLC entropy coding as discussed above, but also will face more serious situations than DCT because that EZW or SPIHT will produce code words with variable length and similarly a single bit of error also will result in the loss of synchronisation. Furthermore, the influence of errors could propagate to the whole frame in video reconstruction.

There have been several approaches to realise the error-resilience of wavelet video coding over error-prone channel. In [127], Creusere proposed the basic idea of realizing the error-resilience by partitioning the wavelet transform coefficients into groups, which will be processed individually by embedded encoder therefore a bit error happened in one group won't affect others. The similar idea can also be found in [128] where the block based coefficient partitioning was proposed.

However, this kind of method that only adding error-resilient format to bit stream at source coder still cannot make the stream robust enough for transmission over error-prone channel. This has lead to the research considering the application of FEC to the video bit stream to against channel errors. In [129-131], the sub bit stream is protected by different code rate of Rate Compatible Punctured Convolutional (RCPC) coding according to the importance of the content of bit stream, thus to achieve the unequal protection. Later Kim *et al.* extended this idea to work with 3D-SPIHT [132], where the ARQ is also adopted which constrains the algorithm to be applied in real time application. In 2002, Cho *et al.* extended the ideas from [127] to 3D-SPIHT and use same RCPC as in [132] to protect resulted packets [133]. The work has been further explored in [134, 135], where the error concealment mechanism and RCPC in conjunction with CRC were added. In [136], Tun *et al.* proposed the similar error-resilient algorithm for DIRAC [137], which is considered as the most mature wavelet video codec so far, by extending the partition wavelet transform coefficients method to motion compensated residual frame and

protecting each packet equally by RCPC or Turbo Code with CRC. These works suggested that FEC has to be adopted in one way or another in order to protect the wavelet-based compressed bit stream. However, the application of FEC inevitably costs huge bandwidth.

The Wyner-Ziv (WZ) coding from distributed source coding theory [3, 38, 40] has been recently adopted as an option for error-resilience in video transmission. It has been shown that using additional WZ bit stream in a systematic lossy error protection framework can provide competitive error-resilience comparing to FEC, especially having advantages over limiting the quality degradation and error propagation with less consumption of bandwidth. In [138, 139], Arron *et al.* first reported the results of applying WZ coding to error-resilient video transmission. The basic idea is that the video signal compressed by MPEG-2 is transmitted over an error-prone channel without any protection. A supplementary stream which is a low rate representation of the transmitted video sequence through coarse quantisation is generated using WZ encoding. The received error-prone MPEG-2 bit stream is used as side information to decode the WZ bit stream. The decoder combines the error-prone side information and the WZ description to yield an improved decoded video signal. The work has been further improved in [140] by composing the WZ stream with coarsely quantised prediction error from MPEG-2 compression and applying RS codes in the WZ codec. The algorithm later was named as Systematic Lossy Error Protection (SLEP) in Rane's later work [141, 142], where further improvements have been proposed including using H.264/AVC to generate the lossy systematic bit stream and using coarse description of redundant slices for WZ bit stream etc. Based on Rane's work in [142], Baccichet introduced the Flexible Macroblock Ordering(FMO) in SLEP to coarsely quantised the region of interest in the frame thus to improve the subject quality [143]. In [144], multiple WZ bit streams containing embedded video descriptions was proposed to better exploit the trade-off between error-resilience and the residual distortion from coarse quantisation in the WZ codec. The similar work can also be found in [145-148].

There is a trade-off between the error-resilience and bandwidth usage. It is always necessary to design such a system to find the balance between them. In many practical applications such as medical image, video surveillance system etc., there are one or more regions of greater interest than others are existed within a frame. Therefore, for this kind of applications, it is unnecessary to waste the bandwidth to protect the whole frame since the end users only concern the quality of the Region Of Interest (ROI) area. Hence it is sensible to give higher priority to those ROI areas than other areas during transmission. Technically speaking, it is feasible to purposely protect the ROI area other than the whole frame. Since only a small bit stream will be generated additionally for ROI area protection comparing to the whole frame protection, huge bandwidth cost can be

saved and the video quality of ROI area can also be improved significantly. By this way the overall system gain can be obtained in terms of compression and bandwidth usage.

In our previous work [149], we proposed a bandwidth efficient error-resilient algorithm for wavelet video coding based on the WZ protection. In the proposed architecture, the video signal is compressed by a generic wavelet video coding to compose the systematic lossy stream and sent through error-prone channel without FEC protection. Meanwhile a supplementary WZ stream which only contains the description of ROI area in wavelet domain will be sent to the decoder for error protection. The ROI area is predefined and the AWGN channel is adopted in this chapter. We further extend the previous work in following aspects: Firstly an automatic ROI detection method is proposed to generate ROI area automatically. Moreover, we use the maximum shift method adopted in JPEG 2000 [150], RDO quantisation [136], and Rate Compatible Punctured Turbo(RCPT) coding [27] to encode the ROI related wavelet coefficients to compose the WZ stream. The generated WZ stream then will be divided into packets for transmission. The packet level LDPC code will be applied to protect WZ packets via adding redundant parity packets. After LDPC decoding, the decoder will combine the received lossy systematic stream and WZ stream to yield the protected ROI related coefficients which promise higher output quality in ROI area, by which the trade-off between bandwidth and error correction ability is optimised. More importantly, in order to satisfy various requirements for heterogeneous groups with various bandwidth conditions and make the scheme compatible in the application like video multicast, the WZER is designed based on a receiver-driven layered protection framework, which enable the receiver choose the best size of WZ stream via joining different layers based on available bandwidth.

5.2 Automatic ROI Detection

ROI area usually is predefined in many applications. However, this is based on the assumption that the encoder has the knowledge of video content thus can manually define the ROI area in advance. In many applications, the video content is not predictable therefore it is hard to define the ROI area in advance. In this paper, an automatic ROI detection method which can be considered as a simple version of video segmentation method of [151, 152] is proposed, by which the subject area with high motion in the frame can be detected automatically and defined as ROI area. This prediction is accurate in most circumstances, especially effective for the video with comparatively static background. The proposed algorithm consists of three processes namely Frame difference mask definition, D_{th} decision and ROI generation.

5.2.1 Frame Difference Mask Definition

Let n denote the length of a group of Frame (GOP), the frame difference between current frame and previous frame is calculated as

$$f_d(x, y, t) = |f(x, y, t+1) - f(x, y, t)| \quad (5.1)$$

where $f(x, y, t)$ is the representation of the frame data in which x, y denote the coordinates and t is frame index with $1 \leq t \leq n-1$. f_d denotes the frame difference. With f_d , the frame difference mask, denoted by f_{dm} can be calculated as:

$$f_{dm}(x, y, t) = \begin{cases} 1, & \text{if } f_d(x, y, t) \geq D_{th} \\ 0, & \text{if } f_d(x, y, t) < D_{th} \end{cases} \quad (5.2)$$

Note that parameter D_{th} need to be set in advance by using $f_d(x, y, 1)$. Exact calculation is shown in Section 5.2.3. Pixels marked by f_{dm} are considered as “moving pixels”.

5.2.2 ROI Generation

According to f_{dm} , pixels moving for a long time are defined as ROI pixels. The procedure of *ROI generation* can be shown as

$$f_n(x, y, n-1) = \sum_{t=1}^{n-1} f_{dm}(x, y, t) \quad (5.3)$$

$$ROI(x, y) = \begin{cases} 1, & \text{if } f_n(x, y, n-1) \geq F_{th} \\ 0, & \text{if } f_n(x, y, n-1) < F_{th} \end{cases} \quad (5.4)$$

where f_n stores the number of times that a pixel moved in the whole GOP and ROI denotes the ROI area. The initial value of f_n and ROI are all set to “0.” The parameter F_{th} is a manually predefined constant which presents the sensitivity of the ROI area detection. A pixel can be viewed as ROI pixel only if it moves for more than F_{th} times.

5.2.3 D_{th} Decision

This section reveals the process to derive the parameter D_{th} used in equation (5.2), which is based on two steps: Gaussianity Test and D_{th} output.

5.2.3.1 Gaussianity Test

Assuming that the frame difference is Gaussian distributed. Thus, Gaussianity test in [151, 152] can be used, which indicates if a group of values is Gaussian distributed or not. Firstly, f_d is divided into many blocks (assume that the size of the block is $M \times N$). Gaussianity test is then applied in each block to examine if the frame difference in the block is distributed in Gaussian or not. The block distributed in Gaussian is deemed to belong to background, and non-Gaussian block belongs to the foreground. The Gaussianity test can be shown as

$$F_{d,k}(t) = \frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N F_d^k(m,n,t) \quad (5.5)$$

$$H(F_{d1}, F_{d2}, F_{d3}, F_{d4}) = F_{d3} + F_{d4} - 3F_{d1}(F_{d2} - F_{d1}^2) - 3F_{d2}^2 - F_{d1}^3 - 2F_{d1}^4 \quad (5.6)$$

where Gaussian is $H(F_{d1}, F_{d2}, F_{d3}, F_{d4}) < G_{th}$ and Non-gaussian: $H(F_{d1}, F_{d2}, F_{d3}, F_{d4}) \geq G_{th}$.

The parameter G_{th} here can be set as a constant “1” because the \mathbf{H} values of foreground blocks and background blocks are dramatically different. The Gaussianity test can roughly distinguish background parts and foreground parts. In order to more precisely distinguish the background and foreground, we need to find the optimum value of D_{th} .

5.2.3.2 D_{th} Output

Since the frame difference is Gaussian distributed, and the digitizing effect of digital systems is considered, the distribution of the absolute value of frame difference in digital domain should be equation (5.7),

$$P_k = \begin{cases} \int_0^{0.5} \frac{2}{\sigma\sqrt{2\pi}} e^{(-x^2/2\sigma^2)} dx, & \text{for } k=0 \\ \int_{k-0.5}^{k+0.5} \frac{2}{\sigma\sqrt{2\pi}} e^{(-x^2/2\sigma^2)} dx, & \text{for } k=1,2,\dots,255 \end{cases} \quad (5.7)$$

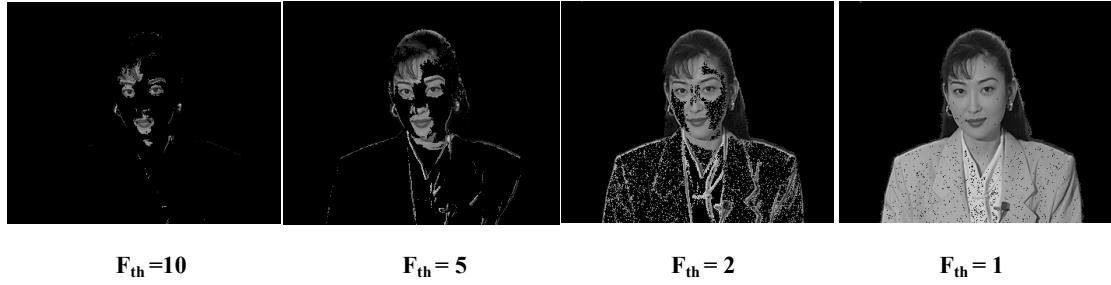
The expected value of the optimal threshold D_{th} should be

$$\begin{aligned} E(D_{th}) &= \sum_{k=0}^{255} k P(D_{th} = k) = \sum_{k=0}^{255} k (P_k)^{F_{th}} \\ P(D_{th} = k) &= (P_k)^{F_{th}} \end{aligned} \quad (5.8)$$

where σ can be calculated instead:

$$\sigma = \sqrt{\sum_{pixels \in backgroundblocks} F_d^2(m,n)} \quad (5.9)$$

Fig.5.1 shows the example of detected ROI area by different values of threshold F_{th} for Akiyo sequence. The value of threshold F_{th} reflects the sensitivity of motion detection. It can be observed in Fig.5.1 that the area with low motion will be neglected during detection if high F_{th} is set and vice versa. On the other hand, F_{th} can be used as a parameter to control the size of ROI area thus generate the different number of ROI related coefficients, which directly determine the size of WZ bit stream later.

Fig.5. 1 ROI area detection with different F_{th}

5.3 ROI Mask Generation in Wavelet Domain

The ROI mask is a bit plane indicating a set of wavelet coefficients whose exact transmission is sufficient for the receiver in order to reconstruct the desired region perfectly. The details of ROI mask derivation should refer to [153]. In this chapter, the mask is a matrix which was initialised to zero with same size of the frame. Following the same steps as the forward transform, the mask is derived by tracing the inverse transform backwards. The positions of all the coefficients used to reconstruct the pixels in ROI area will be marked in mask. However, it should be mentioned here that in order to get the complete reversible transform, the integer wavelet transform based on lifting scheme is used here[154]. In this thesis, the CDF 9/7 wavelet[154] (where 9 and 7 denote the number of filter taps) which is selected as one of the wavelets for the JPEG2000 image format is considered[155].

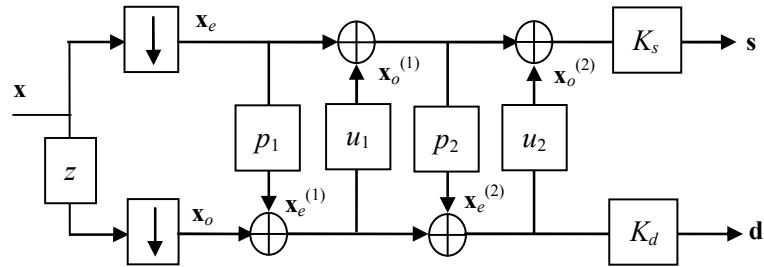


Fig.5. 2 Wavelet transform with lifting scheme

The lifting scheme analysis is described with a sequence of “predict” and “update” filters. Consider a signal $X = (x_k)$ $k \in Z$ with $x_k \in R$. It is first split into two disjoint sets: the even indexed samples $X_e = (x_{2k})$ $k \in Z$ and the odd indexed samples $X_o = (x_{2k+1})$ $k \in Z$. Typically these two sets are closely correlated. Let $p_i(X_e)$ (where $i=1,2$ in CDF9/7 wavelet) denote the predict operation on X_e , $u_i(X_o)$ (where $i=1,2$ in CDF9/7 wavelet) denote the update operation on X_o . The lifting scheme of CDF9/7 is shown in Fig.5.2. In Fig.5.2, p_1, p_2, \dots denote predict filters, and u_1, u_2, \dots denote update filters, $K_s = \zeta$ and $K_d = 1/\zeta$. The parameters for the filters in the format of Z transform are as follows,

$$p_1(z) = \alpha(z + 1), \alpha = -1.586\ 134\ 342\ 060 \quad (5.10)$$

$$u_1(z) = \beta(1+z^{-1}), \beta = -0.052\ 980\ 118\ 573 \quad (5.11)$$

$$p_2(z) = \gamma(z+1), \gamma = 0.882\ 911\ 075\ 531 \quad (5.12)$$

$$u_2(z) = \delta(1+z^{-1}), \delta = 1.149\ 604\ 398\ 860 \quad (5.13)$$

To obtain the detail information d and smoothed values s , we have

$$s = x_e + x_o^{(1)} + x_o^{(2)} \quad (5.14)$$

$$d = x_o + x_e^{(1)} + x_e^{(2)} \quad (5.15)$$

where

$$x_e^{(1)} = p_1(x_e) \quad (5.16)$$

$$x_o^{(1)} = u_1(x_o + x_e^{(1)}) \quad (5.17)$$

$$x_e^{(2)} = p_2(x_e + x_o^{(1)}) \quad (5.18)$$

$$x_o^{(2)} = u_2(x_o + x_e^{(1)} + x_e^{(2)}) \quad (5.19)$$

After the “predict” and “update” filtering steps, X_e is multiplied by K_s and X_o is multiplied by K_d . For the inverse transform, undo the K_s and K_d scale factors, change additions to subtractions, and perform the filtering steps in the reverse order. During the transform, at each decomposition level, the mask will indicate which coefficients are needed exactly at this level so that the inverse transform will reproduce the ROI related coefficients in previous level exactly. The ROI mask matrix is grown slowly following the forward transform until the whole transform is finished. According to this mask, the ROI related coefficients will be picked out and sent to WZ codec for transmission. A typical example definition of ROI mask is show below.

$$M(x,y) = \begin{cases} 0, & \text{the coefficient located at } (x,y) \text{ is not related to ROI} \\ & \text{therefore can be neglected.} \\ 1, & \text{the coefficient located at } (x,y) \text{ is related to ROI and} \\ & \text{need to be protected.} \end{cases}$$

Fig.5.3 shows a generated ROI mask for detected ROI area by $F_{th}=2$. The coefficients highlighted in non-black area are the ROI related coefficients and need to be specially protected to compose the WZ steam.



Fig.5.3 ROI mask example

5.4 Receiver Driven Layered Wyner-Ziv Coding

Wyner-Ziv coding refers to distribute lossy source coding with side information at the decoder. It is suggested in [1, 2] that efficient compression can be achieved if two statistical dependent sources X and Y are separately coded and jointly decoded. In Wyner-Ziv coding, X is coded without knowing of Y , and the decoder conditionally decode X with Y . In practical Wyner-Ziv coding applications, this usually is achieved via following method. The Wyner-Ziv source X is coded by certain type of channel codec (such as turbo code, LDPC code etc), but the systematic bits will be discarded and only the generated parity bits will be sent. The decoder need to first generate the side information Y , which is considered as channel corrupted or estimated version of X . Then by combining Y with received parity bits, X is estimated and decoded at decoder side. In the WZ error-resilient architectures introduced in [138, 139], the side information is extracted from the main systematic system passed through the channel, which can be regarded as the error description of encoded WZ information. Similarly, in this section, the WZ stream contains the description of ROI in wavelet domain. The corresponding side information is located via using the ROI mask for systematic stream. With the side information and received parity bits, the WZ stream can be decoded. The advantages of applying WZ coding algorithm in our error-resilient framework are two folds. Firstly, only the ROI area is protected, hence the size of WZ steam is very small and the bandwidth usage is also small. Secondly, since we don't send the systematic bits after encoding, the number of parity bits sent is based on the coding rate chosen, the bandwidth is saved much again in this way comparing to normal FEC.

Moreover, the proposed error-resilient scheme is based on a receiver-driven layered protection framework, namely, the receiver other than the encoder determines which layer they should join in thus to get the suitable size of WZ stream they can afford according to own available bandwidth. Basically, there are three factors that directly affect the size of WZ stream, which are

F_{th} value, the number of sub bands of ROI related coefficients need to be protected and the encoding rate of turbo codec in WZ codec. According to this, the whole layered structure is designed and shown in Fig. 5.4. Firstly, the first layer is called F_{th} layer, in which F_{th} can be set to determine the size of ROI area. The second layer is called sub band layer, which decide the number of sub bands of ROI related coefficients should be protected via WZ codec. More sub bands of coefficients are contained in WZ stream leads to more details in ROI area recovered. The third layer is called parity layer, in which different coding rates are selected which results different puncturing to generate different size of parity bits set in WZ codec.

Fig.5.4 shows the structure of layer design, the receiver chooses the best option at each layer thus to determine a suitable size of WZ stream to achieve corresponding error-resilience performance.

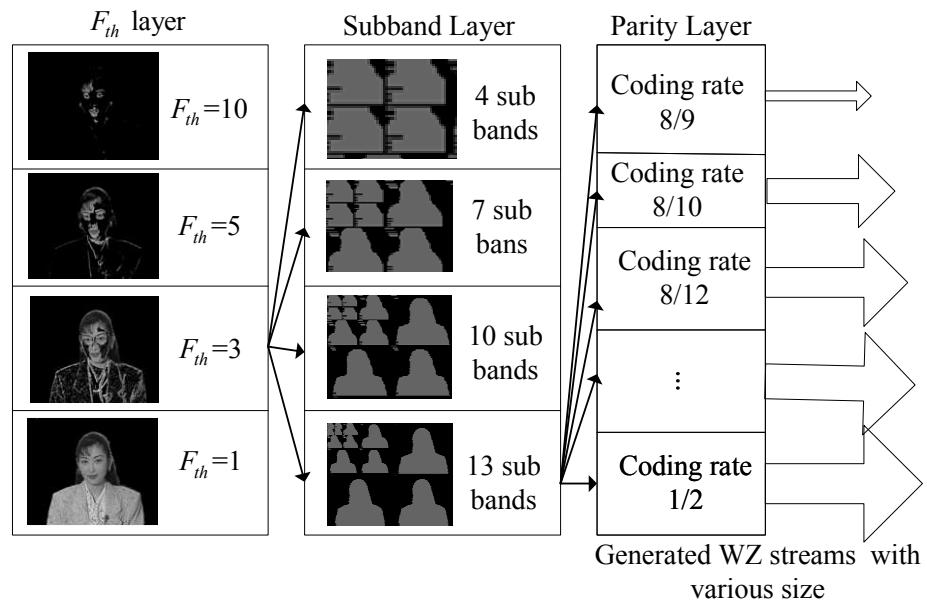


Fig.5. 4 Layer design of the proposed WZER scheme

5.5 WZER for Packet Erasure Channel

In this section, we reveal the WZER performance for wavelet video transmission over packet erasure channel. An additional packet level LDPC codec is required to protect the WZ packets, and then delivered to the decoder with error free.

5.5.1 Architecture

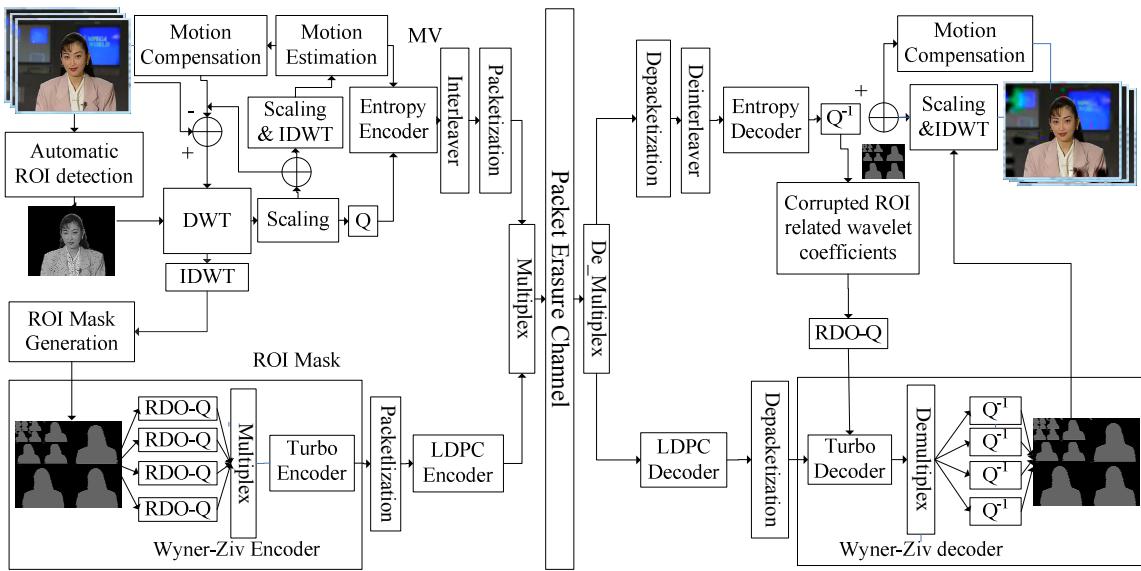


Fig.5.5 The proposed WZER for wavelet video coding over packet erasure channel

Fig.5.5 illustrates the architecture of WZER. The system is composed by two streams, a main wavelet compressed systematic stream and an additional WZ stream to provide error-resilience.

Firstly, the input signal passes through main wavelet video codec and is compressed. The encoder adopts a generic video prediction structure, DWT, and entropy coding to realise temporal, spatial and data compression. Same as in most popular video compression standards, the GOP mode is introduced where there are three types of frames: I, P and B frame where I frame is coded by intra frame coding, P and B frames are coded by inter-frame coding respectively. The fast integer DWT with lifting scheme is applied to perform transforming in order to get the perfect reconstruction in IDWT.

The bit streams from all sub bands are firstly interleaved and divided into packets to transmit over the packet erasure channel without any protection. Meanwhile at the encoder, the video frames of a GOP are sent to a video buffer first, the proposed automatic ROI detection method will be applied to determine the ROI area for current GOP and each GOP will be assigned one ROI area in order to keep ROI area updated during encoding. After the ROI area is defined, a ROI mask is generated after DWT decomposition, in which the positions of ROI related wavelet coefficients are marked. The ROI related coefficients in each sub band will be individually scanned and uniformly quantised. The so called rate distortion optimisation (RDO) quantisation is performed for each sub band, in which the best quantiser for current sub band will be chosen by minimizing the Lagrangian combination of rate and distortion. The generated quantised symbols

from each sub band will be binarised into bit streams, which then are multiplexed into one serial bit stream. The multiplex is performed in the way that the lower frequency sub bands are placed first then follow higher frequency bands. The purpose is to make the stream robust to the channel loss so that if the video bit stream is truncated at any time during transmission, the end user still can use currently received bit stream to realise the partial error-resilience. The multiplexed stream is fed into turbo encoder for encoding, which is implemented by RCPT [27] with the ability to dynamically rate control of WZ stream thus achieve different error protection.

After the encoding, the systematic bits will be discarded and only the parity bits will be sent as WZ stream. The WZ stream will be divided into the packets to transmit. In order to combat with the packet loss in the channel and guarantee the WZ packet delivered with error free, the packet level FEC protection is adopted. Here packet level LDPC code is applied to protect the WZ packet. The WZ packets and the generated LDPC parity packets will be multiplexed with packets from main systematic stream for transmission.

It should be specially mentioned here that for each GOP, only the ROI area of I frame will be protected, with which the corresponding ROI area in P or B frames can be better reconstructed as well since I frame is used as reference frame to build P or B frames. Note that ROI mask and quantisation parameter of each sub band still need to be sent to the decoder via normal channel in order to rebuild the side information. Since the ROI mask only counts for a negligible size of data stream, we assume this mask can be perfectly received at the decoder. In practice, this can be achieved by encapsulating the ROI mask to the header of the main stream etc. At the decoder, the received packets will be regrouped to main systematic packets, WZ packets and LDPC parity packets. The LDPC decoding and the depacketisation process are performed and the main systematic stream and WZ stream are obtained again. At the turbo decoder, by using received ROI mask, the error corrupted wavelet coefficients of ROI region will be picked out from main systematic stream and used as side information to help the turbo decoder to perform decoding. The error-prone wavelet coefficients in same wavelet decomposition level will be quantised by the same procedure as that done at the encoder and play the role of channel corrupted systematic bits to help turbo decoding with received parity bits from WZ stream. After turbo decoding, all symbols of the current level will be dequantised and the ROI related wavelet coefficients will be rebuilt, which later then replaces the error corrupted ROI coefficients in systematic stream. The IDWT is performed afterwards and the whole picture with better ROI reconstruction is rebuilt.

5.5.2 Packet Level LDPC

There have been a lot of researches around the packet level FEC protection in order to against the

packet loss [156-158]. In this thesis, an advanced packet level LDPC-triangle codec [159], which recently is adopted as the standard of Internet Engineering Task Force (IETF), is used to protect the generated WZ stream over packet erasure channel. The LDPC codec used is a newly powerful FEC codec which can survive in the packet loss channel with loss rate nearly up to 30% with code rate of 2/3.

LDPC codes are linear codes obtained from sparse bipartite graphs. Suppose that G is a graph with n left nodes (called message nodes) and r right nodes (called check nodes). The graph gives rise to a linear code of block length n and dimension at least $n-r$ in the following way: The n coordinates of the codewords are associated with the n message nodes. The codewords are those vectors (c_1, \dots, c_n) such that for all check nodes the sum of the neighbouring positions among the message nodes is zero. Fig.5.6 gives an example.

The graph representation is analogous to a matrix representation by looking at the adjacency matrix of the graph: let H be a binary $r \times n$ -matrix in which the entry (i,j) is 1 if and only if the i th check node is connected to the j th message node in the graph. Then the LDPC code defined by the graph is the set of vectors $c = (c_1, \dots, c_n)$ such that $H \cdot c^T = 0$. The matrix H is called a parity check matrix for the code. Conversely, any binary $r \times n$ -matrix gives rise to a bipartite graph between n message and r check nodes, and the code defined as the null space of H is precisely the code associated to this graph. Therefore, any linear code has a representation as a code associated to a bipartite graph (note that this graph is not uniquely defined by the code). However, not every binary linear code has a representation by a sparse bipartite graph. If it does, then the code is called a LDPC code.

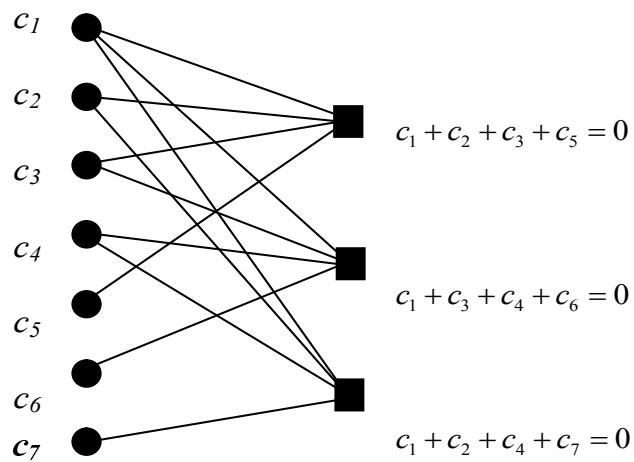


Fig.5. 6 The coding process of LDPC code [160]

The most common algorithms used to decode LDPC codes are the belief propagation algorithm,

the message passing algorithm and the sum-product algorithm. In this thesis we adopt message passing algorithm also called iterative algorithms. One important subclass of message passing algorithms is the belief propagation algorithm. This algorithm is present in Gallager's work [161], and it is also used in the Artificial Intelligence community [162]. The messages passed along the edges in this algorithm are probabilities, or beliefs. More precisely, the message passed from a message node v to a check node c is the probability that v has a certain value given the observed value of that message node, and all the values communicated to v in the prior round from check nodes incident to v other than c . On the other hand, the message passed from c to v is the probability that v has a certain value given all the messages passed to c in the previous round from message nodes other than v .

5.5.3 Results and Discussions

In this section, the performance of WZER is demonstrated.

- Turbo code composed by two identical constituent convolution encoder of rate 1/2 with constraint length of 4 and with polynomial generator of (13,11) is adopted in WZ codec. The puncturing period of RCPT is set to 8, which provided various code rates of (8/9,8/10,8/11,...,1/3 etc) for WZ stream.
- Two CIF sequences “Silent” and “Akiyo” have been investigated during simulation.
- DIRAC [136, 137], is used as wavelet video codec to generate the main systematic stream. In DIRAC, The wavelet transform filters used are the Daubechies (9,7) filter with lifting scheme to perform fast integer wavelet transform for 4 levels, thus total number of sub bands is 13. GOP size is 36 with structure of $L_1L_3L_3L_2L_3L_3L_2$ (DIRAC definition, similar as IBBPBBP structure). The quality factor is set to ‘7’ and the function of coefficient partition is enabled with ‘33’ format [136], with which “Silent” and “Akiyo” are compressed with bit rate of 246.5kbps and 218.1kbps respectively.
- The frame rates for both sequences are 13fps therefore the frame rate actually for WZ codec is only 0.3fps, since the WZER only deal with L_1 frame (I frame). Other parameters for DIRAC keep default value.

The rate of WZER can be varied and influenced by different options in different layers. In following sections, we will show the error-resilient performance of WZER with different rates caused by different combinations of choices of different layers under certain channel condition. For example, we choose $F_{th}=5$ first, protect ROI related coefficients of all 13 sub bands and test the coding rate with 8/14 and 8/16 for WZ stream or we fix the coding rate and number of sub bands need to be protected, but vary the value $F_{th}=1,5,7$ etc.

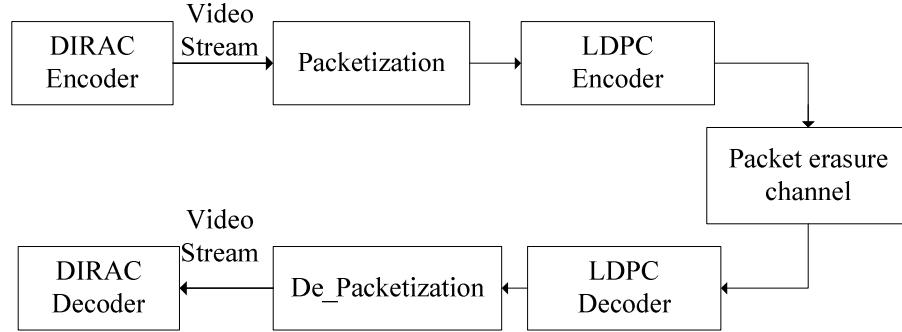


Fig.5. 7 LDPC FEC scheme

In Table 5.1 and Table 5.2, we list suggested schemes for comparison, which are two types of WZER schemes, FEC scheme (LDPC codec with coding rate 8/10 or 8/12 are used, as shown in Fig.5.7) and No protection scheme. Their individual occupied bandwidths also are listed in Table 5.1 and Table 5.2. Further comparisons including average PSNR performance of Y component of the video sequence, and perceptual reconstruction quality and bandwidth usage will be revealed in later sections.

Table 5.1 Error-resilient schemes for Akiyo over packet erasure channel

Table 5.2 Error-resilient schemes for Silent over packet erasure channel

ANC=Average number of ROI related coefficients protected per I frame

ANB=Average number of bits after RDO quantisation per I frame

CR= Coding rate of turbo code

FR= Frame rate

RBS=Resulted bandwidth by scheme.

ABR=Approximate bandwidth caused by LDPC coding (coding rate 8/12) to protect WZ packets in simulation

TAB=Total additional bandwidth required for error-resilience

The formulation to calculate bandwidth usage of WZ stream (namely RBS for WZ) is:

$$RBS_{WZ} = \frac{ANB * \left(\frac{1}{CR} - 1 \right) * FR}{1024} \quad (5.20)$$

N/A=Not applicable

Note: The choice of transformation layer is 13 for all proposed WZER schemes in the table

5.5.3.1 Average PSNR performance

Fig. 5.8 shows the average PSNR performance of Y component of “Silent” and “Akiyo” protected by different protection schemes over packet erasure channel. As shown in Fig. 5.8, the packet drop has corrupted the video stream seriously and very low PSNR is observed for video scheme with no protection.

It can be seen that with FEC schemes, the video stream is protected and PSNR is much enhanced. For both cases, FEC with coding rate of 8/12 can survive in the channel with packet loss up to 30%. FEC with coding rate of 8/10 has inferior performance to that of 8/12 and the PSNR dropped dramatically after 15% packet loss. It even makes the result worse at the point of 30% packet loss, where LDPC decoder cannot handle with the errors happened in the stream and gives more errors during decoding.

With WZER, the additional WZ stream obviously enhances the PSNR to certain extent comparing to no protection. Higher the rate of WZ stream is, the better PSNR gain can be obtained. It should be noticed that the WZ stream occupies much less bandwidth comparing to FEC protection. However, it is apparent that the average PSNR of FEC is higher than that of WZER in Fig.5.9. But does higher PSNR means FEC performs better than WZER? PSNR is not the only criteria to evaluate the reconstruction quality, and next section will give more answers to this question by revealing the perceptual performance of these schemes.

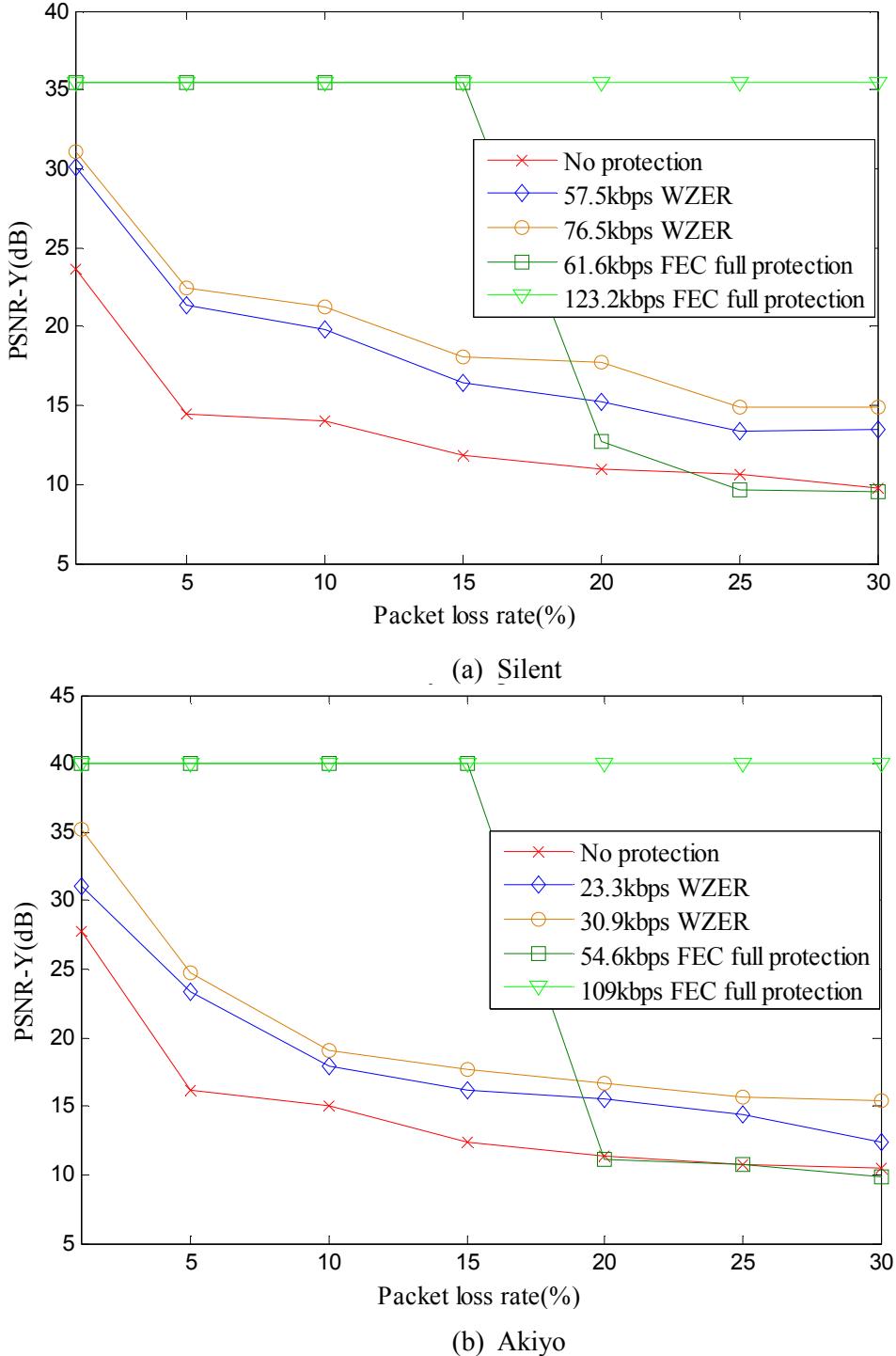


Fig.5.8 Average PSNR performance of WZER under packet erasure channel

(13 sub bands protected, Packet loss rate =1%~30%)

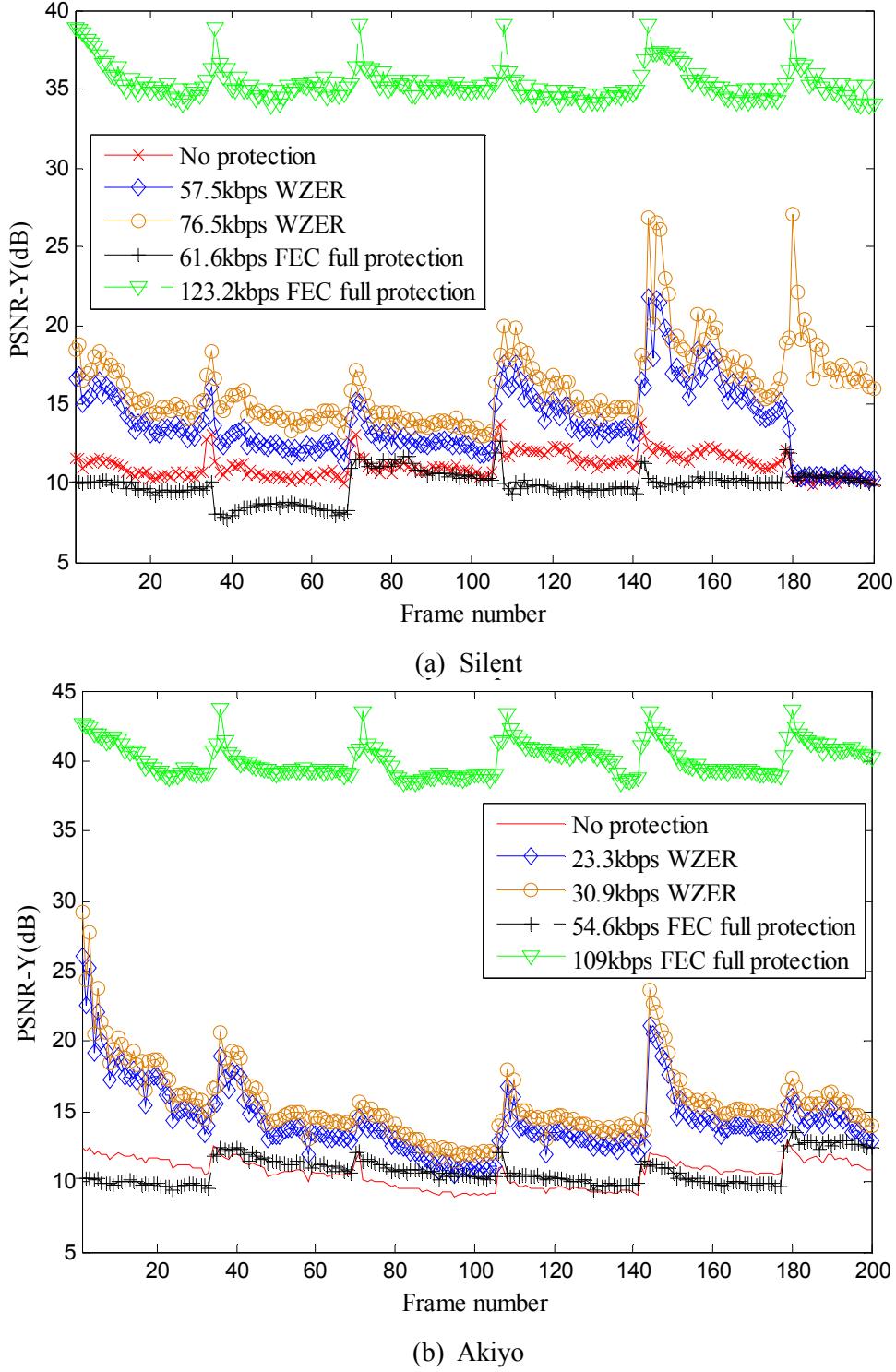


Fig.5.9 PSNR performance vs frame number of WZER

(Packet loss rate =25%; 13 sub bands protected; frame number= 1~200)

5.5.3.2 Perceptual performance comparison

Sometimes PSNR cannot truly reflect the true quality of video protected, we illustrate the

perceptual comparison of L_1, L_2 and L_3 , namely I, P and B frames of “Akiyo” sequence in Fig.5.10. The perceptual comparison has clearly showed the advantage of WZER over FEC scheme.

It can be found in Fig. 5.10 that 54.6kbps FEC still cannot fully correct errors happened in the frame at 25% packet loss. With FEC, many errors happened at background are corrected but the ROI area still cannot be recovered. This is not acceptable for end users who actually concern the quality of ROI area much more than the rest areas. In other words, the redundant parity bits sent in FEC actually are wasting in protecting those background areas, which are not interest of the end users. In order to correct more errors happened in ROI area, more redundant bits are required. In this case, 109kbps FEC would be enough to successfully decode all the errors in the frame. However, larger bandwidth is inevitably required.



Fig.5. 10 Perceptual comparison of I, B, P frames with WZER of Akiyo sequence

($F_h=5$; 13 sub bands protected, Packet loss rate =25%)

Conversely, 30.9kbps WZER has fully utilised parity bits to protect the ROI area thus has higher quality of ROI area than that of 54.6kbps FEC. Especially the most important part in the

frame is perfectly protected. In Fig. 5.10(a), I frame is partially protected and 30.9kbps WZER makes comparable quality in ROI area virtually with 109kbps FEC. Furthermore, it can be observed that not only the ROI area has significant improvement, but also the adjacent area of ROI area has also been improved to certain extent. This is due to that some protected ROI related wavelet coefficients are not only used for reconstructing the ROI area, but also used by some other pixels located in the adjacent area for IDWT reconstruction. Even though WZER only works for I frame, it also can be found that in Fig. 5.10(b) and (c), the output quality of ROI area in P and B frames are correspondingly improved following the enhancement of I frame. Because I frame is the first basic reference frame to reconstruct the following P and B frames, the improvement of I frame has been propagated to the whole GOP. Not only clear perceptual improvement can be observed in Fig.5.10 (b) and (c), Fig. 5.9(a) and (b) further proved the trend that each enhancement in I frame could lead to the corresponding improvement in whole GOP. However, some white dots error still cannot be eliminated inside ROI area of P and B frame in Fig.5.10. These errors are caused by the errors of MV and residual, which are not covered by WZER thus cannot be corrected.

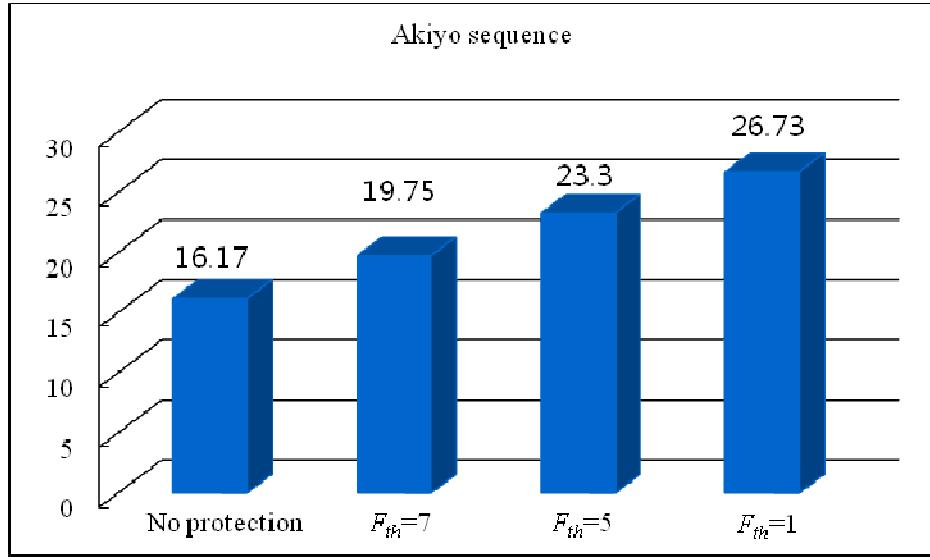
5.5.3.3 Bandwidth utilisation analysis

In the proposed applications, the perceptual quality is more concerned than the PSNR. The PSNR gain can be dropped because of bad quality of areas outside the ROI, which are not necessary to be protected. However, the traditional FEC algorithm cannot distinguish which part of the bit stream should be protected therefore a part of bandwidth actually is wasted for the protection of unnecessary area. For the proposed WZER scheme, the most important area (ROI) has been picked out and the limited bandwidth have been efficiently utilised to protect it. The significant improvement in ROI area of the frame can be observed in Fig.5.10. Therefore, given the condition that the end user only concerns the quality of the ROI area in the frame, the WZER scheme actually outperforms FEC scheme with less bandwidth. Take Akiyo's case as an example, 30.9kbps WZER actually gives more satisfying output than 54.6kbps FEC and brings the same effect as 109kbps FEC to end users, but only occupies around 28.3% of bandwidth. Higher F_{th} could lead to further reduction in bandwidth while offer different error-resilient performance, which is revealed in the next section.

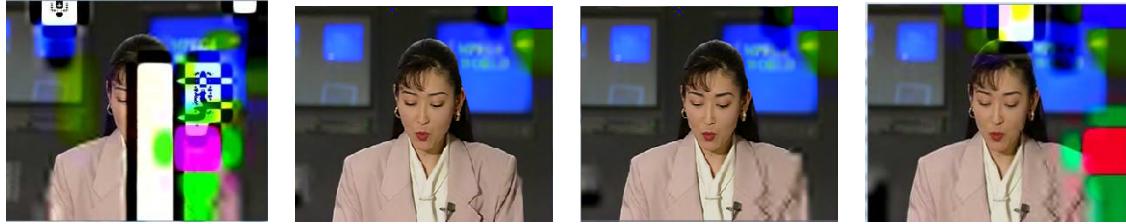
5.5.3.4 WZER performance with different F_{th}

Fig. 5.11 shows the Y component PSNR performance and perceptual quality of Akiyo sequence protected by WZER with different F_{th} values. As analyzed in previous sections, F_{th} value directly controls the size of ROI area, which results in the different size of WZ stream eventually. Fig.5.11

(a) shows the average PSNR gain of different WZ stream. Smaller F_{th} value gives larger bit rate of the WZ stream, and hence higher PSNR gain can be expected. In Fig. 5.11 (b), it can be observed that higher F_{th} results in smaller ROI area required to be protected. In the case of $F_{th}=1$, the basic shape Akiyo has been nicely protected. When it changes to $F_{th}=5$, some blur areas can be spotted around the right shoulder part. The protected ROI area is shrunk into only the face part of Akiyo for the case of $F_{th}=7$.



(a) Average PSNR comparison



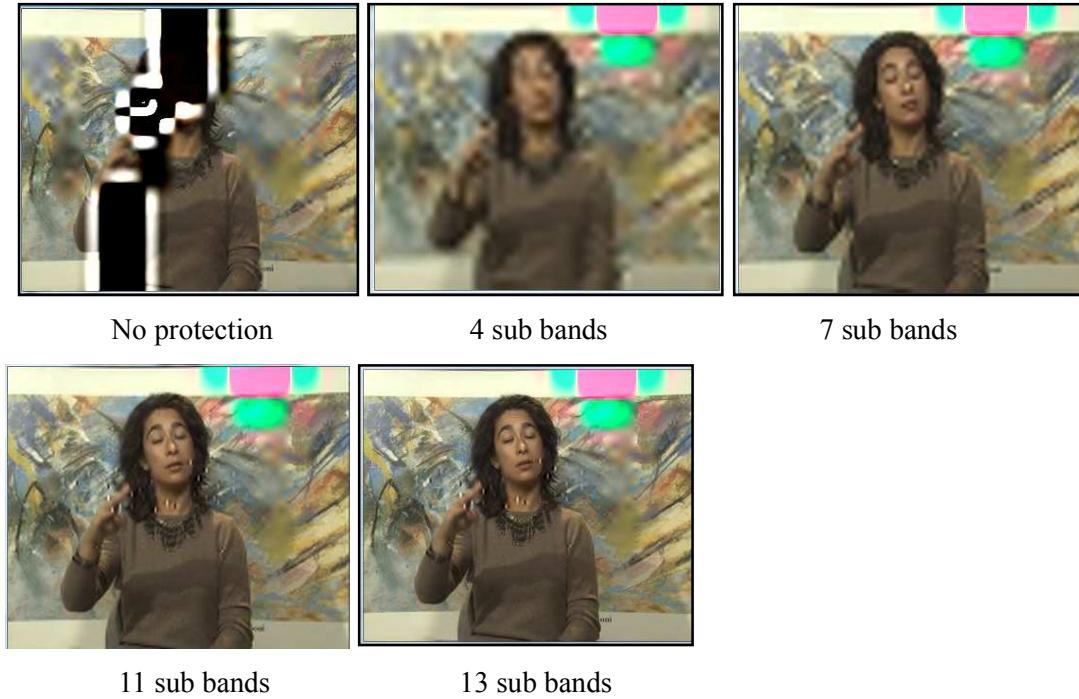
(b) Perceptual comparison

Fig.5. 11 PSNR and perceptual performance for WZER with different F_{th} (WZ Coding rate=8/14, 13 sub bands protected, Packet loss rate 5%, $F_{th}=1,5,7$, Frame: 108).

5.5.3.5 WZER performance with different number of sub bands protected

The number of sub bands of ROI related coefficients is another key factor that can influence the size of WZ stream. More sub bands sent would generate higher bit rate of WZ stream but more details can be protected and better recovered and vice versa. In the proposed layered framework, the end user can choose the number of sub bands that they want to receive judged by the resolution level they are satisfied with. For example, as shown in Fig.5.12, 7 sub bands approximately can satisfy the requirement of application like video surveillance system etc. But

for application in medical image etc, more than 10 sub bands probably are required. Moreover, the WZ stream is composed in the way that stream from lower frequency sub band is placed ahead in the stream. This gives the algorithm advantage to combat with the channel loss. If the WZ stream is truncated during the transmission, the decoder can use the currently received sub bands to partially recover ROI area with its best.



**Fig.5. 12 Perceptual comparison for WZER with different number of sub bands received
(Packet loss rate=5%, WZER coding rate= 8/14 ; $F_{th}=3$, the 36th Frame)**

5.6 WZER for AWGN Channel

In this section, we reveal the WZER performance for wavelet video transmission over additional white Gaussian noise (AWGN) channel.

5.6.1 Architecture

Fig.5.13 illustrates the WZER for wavelet video transmission over AWGN channel. The system is composed by two channels, a main wavelet compression channel and an additional WZ channel to provide error resilience.

Firstly, the input video sequence passed through a generic wavelet based video codec to generate the main systematic stream. The wavelet codec is implemented by a generic video prediction structure, in which DWT, ME and entropy coding used to realise temporal, spatial and data compression. The GOP mode is also introduced in which I, P and B frame are used. Typically, I frame is coded by intra-frame coding while P frame and B frame are coded by inter-frame

coding. The fast integer DWT with lifting scheme is applied to perform transforming in order to get the perfect reconstruction in IDWT. The generated bit stream which is referred as main stream is passed through the error prone channel without any FEC protection.

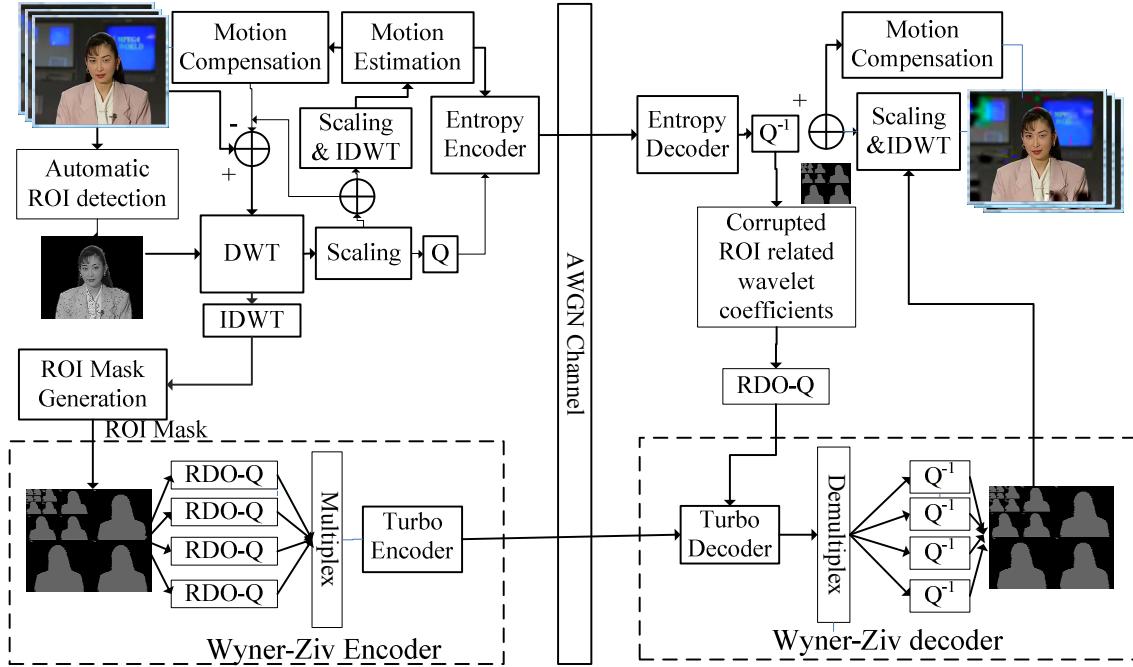


Fig.5. 13 The proposed WZER for wavelet video coding over AWGN channel

Meanwhile at the encoder, the video frames of a GOP are sent to a video buffer first, the proposed fast video segmentation method will be applied to automatically detect the ROI area for the current GOP and each GOP will only be assigned one ROI area. After the ROI area is defined, a ROI mask for reconstructing the ROI is generated after DWT decomposition, in which the positions of ROI related wavelet coefficients are marked. The ROI related coefficients in each sub band will be individually scanned and uniformly quantised. The RDO quantisation is performed for each sub band, in which the best quantiser for the current sub band will be chosen by minimizing the Lagrangian combination of rate and distortion. The generated quantised symbols will be binarised into bit streams. The whole bit streams is composed by concatenating all sub band bits streams together following the order from the lowest frequency band to the highest. This operation can make the stream robust to the channel loss so that if the bit stream of video is truncated at any time during transmission, the end user still can use currently received bit streams to realise the partial error resilience.

The bit stream is then sent into the turbo encoder for encoding. Again, the RCPT is used here to realise dynamic rate control of WZ stream thus offers different error protection. Note that only the generated parity bits will be used to compose the WZ stream while the systematic bits are

discarded. The length of WZ stream namely number of parity bits needed to be sent depends on the coding rate chosen. WZER only protect the ROI area of I frame, with which the corresponding ROI area in P or B frames can also be better reconstructed with better quality since I frame is used as reference frame to rebuild P or B frame. Note that ROI mask and quantisation parameters of each sub band need to be sent to decoder via normal channel in order to rebuild the side information. Since the ROI mask only count for a negligible size of data stream, we assume this mask can be perfectly received at the decoder. In practice, this can be achieved by encapsulating them to the header of the main stream etc.

At the decoder side, the wavelet video decoder will try to rebuild every sub band in wavelet domain based on the received main systematic stream. By using ROI mask, the error corrupted wavelet coefficients of ROI region will be picked out and used as side information to help the turbo decoder to decode the WZ stream. During the decoding, the error-prone wavelet coefficients in same wavelet decomposition level will be quantised by same procedure as that done at the encoder and play the role of channel corrupted systematic bits. After turbo decoding, all symbols for the current level will be dequantised and the ROI related wavelet coefficients will be rebuilt, which later then replace the error corrupted ROI coefficients. The IDWT is performed afterwards and the whole picture is rebuilt, after which the ROI area will be reconstructed with higher quality than the rest areas.

5.6.2 Simulation Results

In this section, we investigate the performance of WZER for wavelet based video transmission over wireless link. Typically, the AWGN channel has been adopted as error-prone channel. DIRAC [136, 137] is used as the wavelet video codec to generate the main systematic stream. The wavelet transform filters used are the Daubechies (9,7) filter with lifting scheme to perform 4 levels wavelet decomposition. Turbo code composed by two identical constituent convolution encoder of rate 1/2 with constraint length of 4 and with polynomial generator of (13,11) is adopted in WZ codec. The puncturing period of RCPT is set to 8, which provides various code rates of (8/9,8/10,8/11,...,1/3) for WZ stream. The frame rate is set to 13fps, GOP size is 36 with structure of IBBPBBP.... Note that since the proposed scheme only deals with I frame, therefore the frame rate in WZ codec is only 0.3fps. Two CIF sequences “Silent” and ”Akiyo” have been tested during the simulation. The proposed WZER is mainly compared to normal FEC scheme which is implemented by using turbo codec with coding rate 8/10 and 8/12.

Table 5.3 and Table 5.4 lists the bandwidth usage for WZER and turbo FEC to deliver “Akiyo” and “Silent” sequences over the AWGN channel with SNR=6dB. The original size of the

compressed streams is 218.1kpbs for Akiyo and 246.5kbps for Silent by using DIRAC. The abbreviation in the tables has same definition as in Table 5.1 and Table 5.2. Note since the architecture is based on the AWGN channel, there is no need to add additional protection to the WZ stream.

Table 5. 3 Error-resilient schemes for Akiyo over AWGN channel

Scheme type	ANC	ANB	CR	FR(fps)	RBS	TAB
No Protection	N/A	N/A	N/A	13	218.1	0
WZER Fth=1	47388	246417	8/12	0.3	36.1	36.1
			8/14	0.3	54.1	54.1
WZER Fth=5	12805	70428	8/12	0.3	10.3	10.3
			8/14	0.3	15.5	15.5
WZER Fth=7	10290	53508	8/12	0.3	7.8	7.8
			8/14	0.3	11.8	11.8
FEC	N/A	N/A	N/A	8/10	13	272.7
			N/A	8/12	13	327.1
						109

Table 5. 4 Error-resilient schemes for Silent over AWGN channel

	ANC	ANB	CR	FR(fps)	RBS	TAB
No Protection	N/A	N/A	N/A	13	246.5	0
WZER Fth=3	34196	174401	8/12	0.3	25.5	25.5
			8/14	0.3	38.3	38.3
FEC	N/A	N/A	8/10	13	308.1	61.6
			8/12	13	368.8	123.2
			8/14	13	431.75	184.75

5.6.2.1 Analysis of PSNR performance and perceptual performance

I) PSNR performance comparison

(For Akiyo $F_{th}=5$, Silent $F_{th}=3$, all 13 sub bands protected and channel SNR=6dB)

Fig. 5.14 and Fig. 5.15 show the average PSNR performance of Y component of Akiyo sequence and perceptual comparison of I, P and B frames of Akiyo sequence protected by different protection schemes.

In Fig.5.14 and Fig.5.15, the similar trend as in packet loss channel can be found. Firstly, without any protection, the video stream is vulnerable to the channel noise and the reconstructed

video quality is very poor. After applying FEC, the video stream is more robust for transmission. 109kbps FEC full protection basically can guarantee the stream delivered to the decoder with error free under such channel condition. The 54.6kbps FEC has inferior performance thus provide lower PSNR. Although the gained PSNR is higher but huge bandwidth consumption cannot be avoided. With WZER, the improvement in ROI area also brings the overall PSNR enhancement

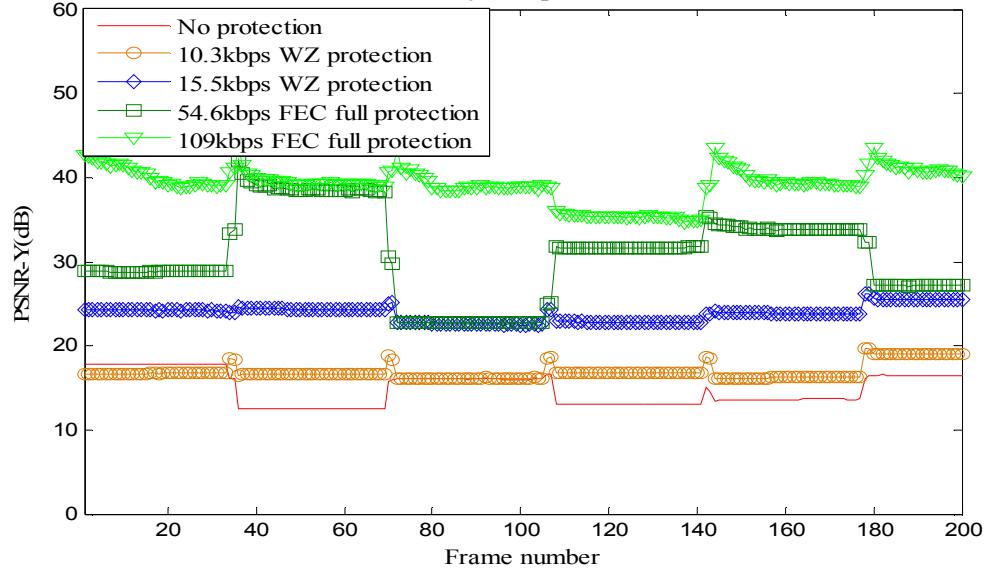


Fig.5.14 PSNR performance of WZER for Akiyo sequence

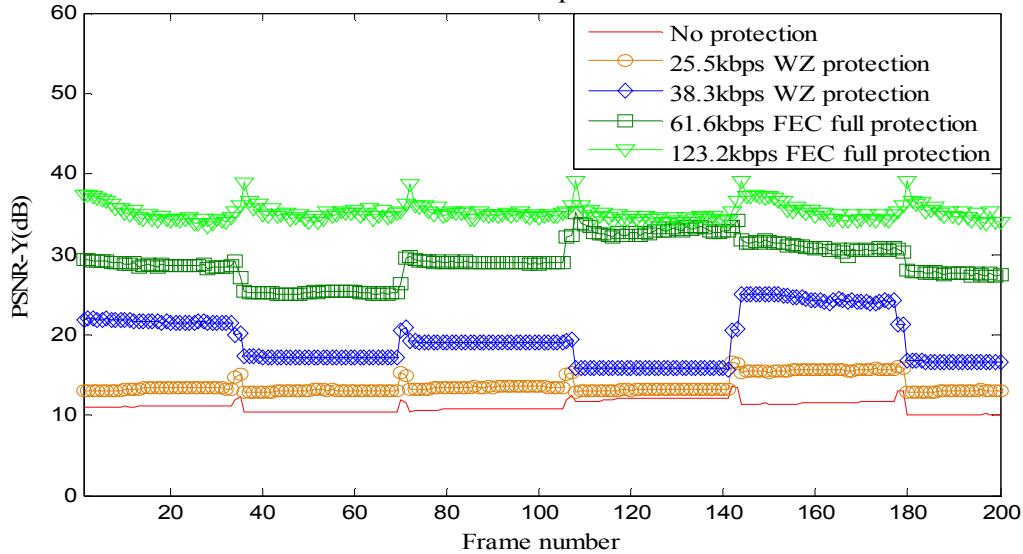


Fig.5.15 PSNR performance of WZER for Silent sequence

to certain extent. The 15.5kbps WZ streams can improve PSNR up to 11dB for Akiyo sequence and 38.3kbps improved 17dB for Silent sequence. The 15.5kbps and 38.3kbps are the highest bit streams that can almost decode the ROI related wavelet coefficients with error free under this channel condition in the simulation, which also implies the largest bandwidth needed for WZER

in the case of example sequences.

II) Perceptual comparison



Fig.5. 16 Perceptual comparison of I, P, B frames of WZER for Akiyo sequence

Fig. 5.16 and Fig.5.17 show the perceptual comparison of I, P and B frames of Akiyo and Silent sequence. In this aspect, WZER performs similar as it works for packet erasure channel. Fig.5.16 (a) shows that I frame protected by 15.5kbps WZER provides the same output quality in ROI area as 109kbps FEC. Furthermore, it also can be observed the improvement in the ROI area leads to the quality enhancement in the adjacent area. Because the recovered ROI related wavelet coefficients are not only used for reconstructing the ROI area, but also used for reconstructing some other pixels located in adjacent area. Fig.5.16 (b) and Fig.5.16 (c) also shows that the output quality of ROI area in P and B frames are correspondingly improved. I frame is the first basic reference frame to reconstruct the following P and B frames thus the improvement of I frame can lead to the overall enhancement in whole GOP. This trend can be further proved in Fig. 5.14 and Fig.5.15 which indicate that the PSRN of P and B frame are improved following every peak of I

frame.



(i)No protection (ii)38.3kbps WZ stream (iii)61.6kbps full FEC (iv)123.2kbps full FEC

(a) Frame36(Iframe)



(i)No protection (ii)38.3kbps WZ stream (iii)61.6kbps full FEC (iv)123.2kbps full FEC

(b) Frame 39 (P frame)



(i)No protection (ii)38.3kbps WZ stream (iii)61.6kbps full FEC (iv)123.2kbps full FEC

(c) Frame 38 (B frame)

Fig.5. 17 Perceptual comparison of I, B, P frames of WZER for Silent sequence

The advantage of WZER over FEC still can be observed by perceptual comparison in this case. Take Akiyo as an example, 15.5kbps can offer better ROI area output than 54.6kbps FEC and virtually same effect as 109kbps FEC. The bandwidth that WZER consumed is much less than FEC but give better ROI output. As introduced in previous simulation part, the WZER can fully utilises the bandwidth to transmit all necessary bits to protect the ROI area while FEC cannot distinguish the ROI part and wastes many bits to protect many unimportant area which are not necessary to be protected for end users.

5.6.2.2 WZER performance with different F_{th}

(Akiyo sequence, Coding rate=8/14, 4 levels, $F_{th}=1,3,5,7$, channel SNR=6dB)

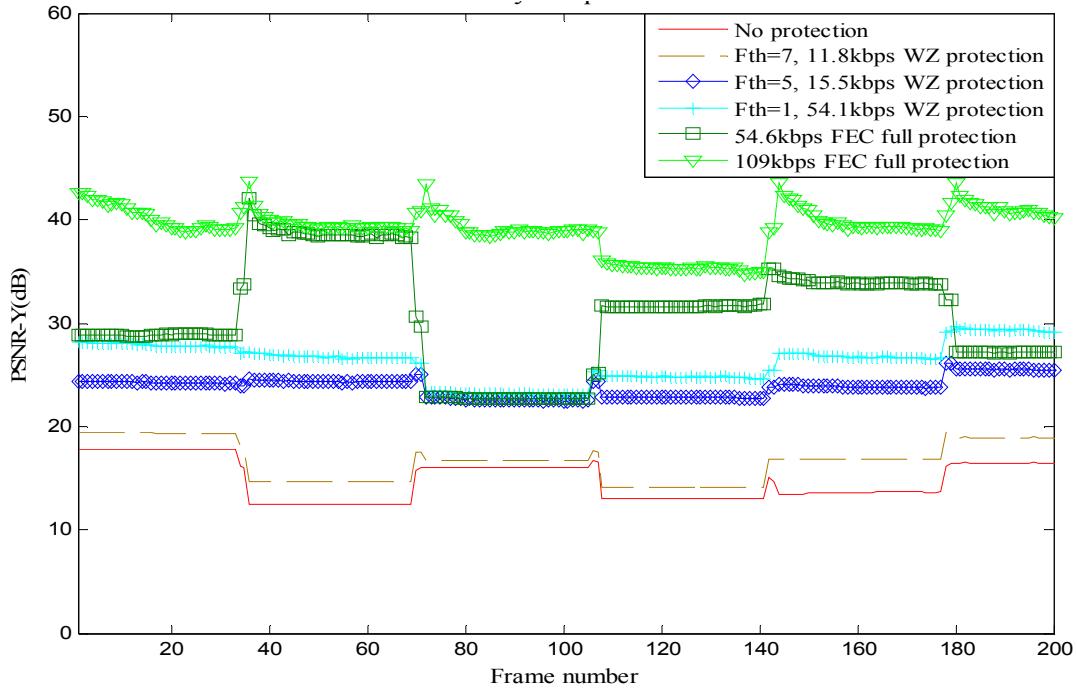


Fig.5. 18 Average PSNR performance of WZER with different F_{th} of Akiyo sequence

Fig. 5.18, Fig. 5.19 and Fig. 5.20 show the PSNR performance of Y component and perceptual comparison of Akiyo sequence protected by WZER with different F_{th} values

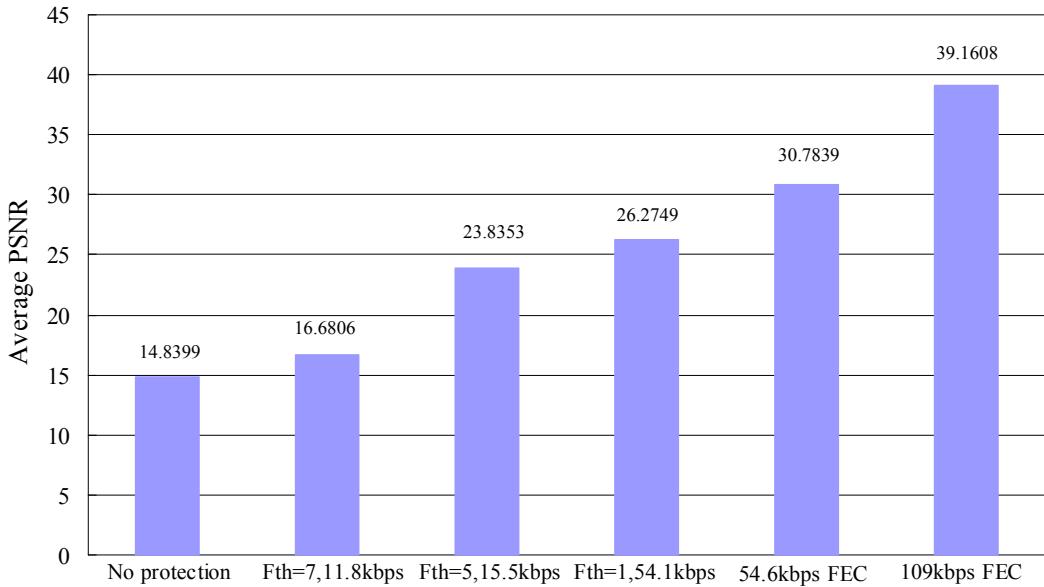


Fig.5. 19 Average PSNR performance of WZ stream with different F_{th} values of Akiyo sequence

F_{th} is a parameter that plays key role in the proposed layer design framework. F_{th} directly control the sensitivity of the ROI detection. Higher F_{th} results a smaller ROI area and vice versa. The size of the ROI area determines the size of the WZ stream. In Fig.5.18 and Fig.5.19, it shows that lower F_{th} generated larger size of WZ stream and leads to higher PSNR performance. In the

perceptual comparison in Fig.5.20, it is clearly seen that lower F_{th} provide larger ROI area protection, but has to occupy higher bandwidth.

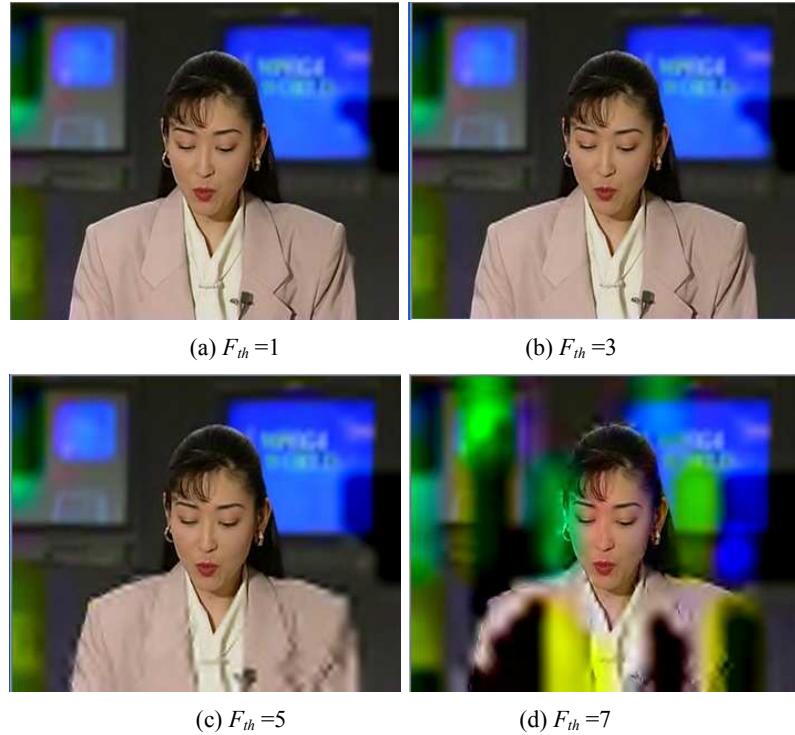


Fig.5. 20 Perceptual quality comparison of WZER with different F_{th} of Akiyo sequence

5.6.2.3 WZER performance with different number of sub bands protected.

Akiyo sequence, $F_{th}=1$, Levels(1,2,3,4), code rate=8/14

Data truncation is usually happened during the transmission, especially in wireless transmission. The WZ stream is composed in the order from low frequency level to high frequency level. This kind of composition can give partial robust to the WZ stream to combat the channel loss. If the WZ stream is truncated during the transmission, the decoder can use the currently received bit streams to partially recover the error resilience. In the simulation, 4 levels wavelet decomposition is performed thus there are total 13 sub bands. Fig. 5.21 shows the perceptual quality of different number of sub bands of the WZ stream received. It can be observed video quality is gradually increased with the more sub bands of bit stream are received. However, if the truncation is happened, the current received bit streams can help to protect the ROI area to certain content.

This design also offers another choice in the layered framework. The end user can choose the number of levels according to available bandwidth of the resolution level requirement of the application.

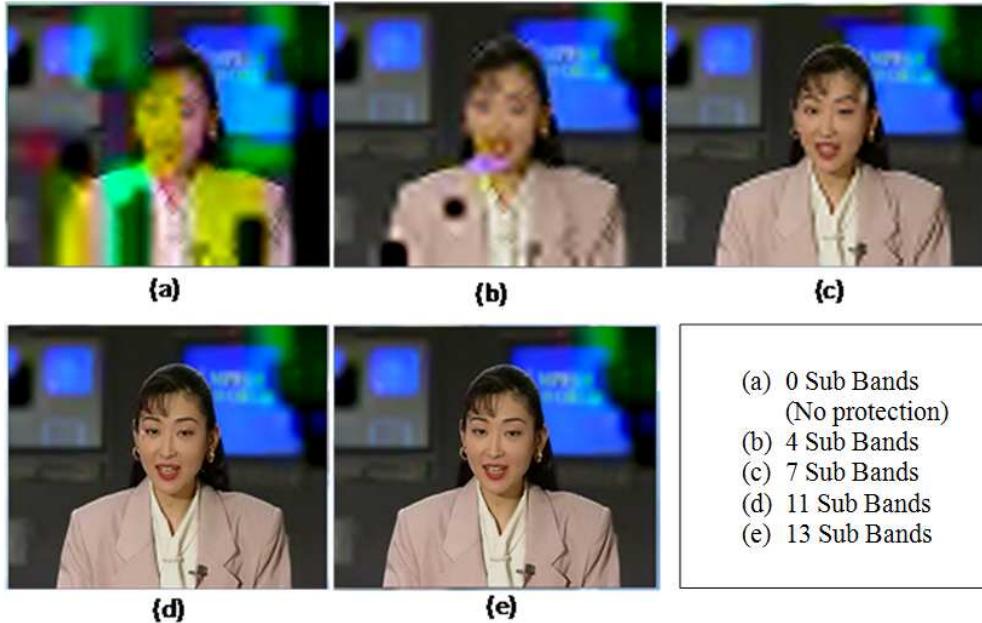


Fig.5. 21 Perceptual comparison for WZER with different number of sub bands of Akiyo sequence

5.6.2.4 Bandwidth analysis

The advantage of the proposed WZER in using the limited bandwidth over FEC is even more evident than it works for packet erasure channel due the lack of the packet level FEC protection to the WZ stream which consumed part of the bandwidth. For Akiyo sequence, 15.5kpbs WZER which is resulted by $F_{th}=5$, coding rate=8/14 and 13 sub bands protected can perform better than 54.6kpbs FEC which is resulted by coding rate of 8/12 in term of final reconstruction quality ROI that end users concern. It also performs same as 109kpbs FEC which is resulted by coding rate of 8/14. In the aspect of ROI area reconstruction, the bandwidth of WZER saved nearly 86% in this case. For Silent sequence, the bandwidth of 38.3kpbs WZER can save up to 69% comparing to the 123.2 kbps full FEC. The size of the WZ stream can be varied by choosing different F_{th} value, different coding rate and different number of sub bands need to be protected. The bandwidth for WZER can be further reduced by using higher F_{th} , higher coding rate and less number of sub bands.

5.7 Chapter Summary

In this chapter, an efficient error-resilient scheme called WZER based on a receiver driven layered WZ coding framework is proposed for wavelet video transmission over error prone channel. The proposed WZER emphasise the protection of the ROI area in the frame, which is believed to be more important than the rest areas and more interesting to end users for some applications. The principle of WZER is to use the limited bandwidth to fully protect the ROI area

while neglecting the quality loss outside the ROI area which is not interest to the end users. With WZER, all the ROI related wavelet coefficients in wavelet domain are protected by WZ coding, in which automatic ROI detection, maximum shift method, RDO quantisation, turbo coding are used separately. By using the error-prone wavelet coefficients from the systematic stream which is original compressed stream from wavelet video codec and sent to the channel without any protection, the protected ROI related coefficients can be recovered at the decoder, with which the final output will be improved with a better reconstruction in ROI area.

In the simulation, the performance of the proposed WZER is verified by simulating it on two types of error-prone channels: packet erasure channel and AWGN channel. In the simulation of packet erasure channel, an additional packet level LDPC codec is used to protect the generated WZ stream. The simulation results reveal that WZER improves the overall system PSNR to certain extent. Although the WZER only works with I frame, it is proved in the simulation that the improvement of PSNR is propagated to the whole GOP. Most importantly, the ROI area has been specially protected and the satisfactory perceptual quality of subject in the frame is obtained in the simulation under harsh channel condition without consuming too much bandwidth. Furthermore, because of the perfect reconstruction of ROI related coefficients, the adjacent areas also are positively affected since they use some of the ROI related coefficients for reconstruction. The receiver driven framework makes the WZER applicable in the application of multicast, where receivers from heterogeneous group with various bandwidth availability can be satisfied. The multiplexing way of WZ stream makes the scheme more robust to the channel loss, in which the decoder can use current received stream to realise partial error-resilience if the truncation is happened to the bit stream during transmission. Comparing to the traditional FEC, the WZER scheme has several advantages over FEC in the aspects of bandwidth efficiency, subject area protection and trade-off between the bandwidth and error-resilience etc.

In our future work, the protection regarding to the MVs and residual related to the ROI blocks in WZER will be discussed and the performance of WZER under channel with higher packet loss rate will be explored.

Chapter 6

6. Conclusions and Future Work

In this chapter, all the research works presented in this thesis will be summarised and concluded along with their overall achievements. After that, all the possible modifications which could improve the performance of the presented research methodologies will be discussed in detail as future works.

6.1 Overall Achievements

6.1.1 Pixel domain WZ Video Coding with Optimised Decoder

In this part of research, four proposals regarding to the optimisation of the decoder in WZ video coding are presented. As introduced in previous chapter, most current WZ video coding issues are aiming to improve the system performance as a whole neglecting the huge complexity caused for the decoder. Although in DVC, the decoder is designed to be complex comparing to the encoder, the complexity of decoder in WZ video coding essentially influences the system coding efficiency. For example, the system is not able to provide timely output due to the delay caused by the complex algorithm at the decoder. Hence it is desirable to design the algorithm to optimise the decoder. The algorithm should reduce the complexity of the decoder without loss of the performance meanwhile. In this thesis, four algorithms are proposed to optimise the decoder in following aspects: optimising the Slepian-Wolf codec, optimising the side information generation, optimising the side information refinement and optimising the feedback process.

In Section 3.1, a slice structure is proposed to reduce the input block size of the Slepian-Wolf codec. The design of the slice structure is based on the practical channel codec selected. With slice structure, it is found that the utilisation of slice structure in WZ video coding brings two advantages. On the one hand, the input block size is reduced and compatible with the channel codec used, the coding efficiency is enhanced significantly comparing to an arbitrary size input block in previous available research issues. On the other hand, slice structure optimises the rate control process adopted in most WZ video coding solutions. With slice structure, the encoder avoids sending unnecessary parity bits thus the R-D performance is improved as a whole.

Moreover, to get accurate side information, complex side information generation methods are proposed in most literatures. It is proved that most side information interpolation methods are not reliable since most of them are based on the assumption of linear translation model of the object inside frame. Based on this fact, it is not necessary to spend much computation in this part. We

propose a fast side information interpolation method called BMRF by using multiple reference frames to provide relatively better side information comparing to the normal side information interpolation method in Section 3.2. The complexity of the side information generation is much reduced.

After that, an embedded refinement technique is proposed to refine the side information in Section 3.3. With the proposed refinement technique, the refinement process is moved into the turbo decoder and operated followed by each iteration decoding of the turbo codec. The advantages of the embedded refinement are two folds. Firstly, enormous repeat decoding process is avoided due to moving the refinement process inside the decoder. Secondly, the system performance loss caused by the side information generated by BMRF can be compensated by the refinement process. Combining the BMRF and the embedded refinement, the decoder is optimised without losing system performance.

Finally, an adaptive feedback optimisation algorithm is proposed in Section 3.4, which is targeting to optimise the feedback process in WZ video coding. The proposed algorithm utilises the feedback channel to send back the evaluation of side information quality, with which the encoder sends the possible number of parity bits that can successfully finish the decoding. The retransmission and repeat decoding can be largely reduced thus the decoding complexity is optimised correspondingly. With above four proposals, an advanced pixel domain Wyner-Ziv video coding with optimised decoder is achieved.

6.1.2 3D-TDWZ and its Application in Multiview Video Coding

Although using transform domain in WZ video coding will slightly increase the complexity at the encoder, the superior performance to the pixel domain WZ video coding can be provided due to exploiting spatial redundancy. The transform domain WZ video coding has same complexity of the encoder as the conventional intra-frame coding standard. However, due to the limitation in the computation complexity, the traditional method of ME used to exploit temporal redundancy is not suitable for WZ video coding. In Chapter 4, a 3D DCT transformation method is proposed for WZ video coding to exploit the temporal redundancy. 3D DCT has several advantages over traditional method in the circumstance of dealing with non-translational motions thus can be used to replace the ME. The transform domain WZ video coding operates similar as previous transform domain WZ video coding based on 2D DCT, the only additional operation is the third 1D DCT in the temporal direction. The WZ frames are processed group by group other than frame by frame which is done in previous transform domain WZ coding solutions. The final simulation results proved that the performance of WZ video coding can be further improved

especially when dealing with video sequences that have high temporal correlation. The computation complexity of this operation is slightly increased but still can be considered as low complexity since 3D DCT is already used in some applications such as wireless sensor network, mobile camera etc.

WZ video coding can be applied in many practical areas and one important application of them is multiview video coding. There have been some attempts to the applications of WZ video coding in last few years and most of them are based on pixel domain. One of the great advantages to apply WZ video coding principle to multiview video coding is that the cameras are low complexity and not necessary to communicate with each other. In Chapter 4, we apply the transform domain WZ video coding into multiview video coding. Particularly, we investigate the performance of three types of transform domain DMVC (transform domain DMVC with 3D DCT, 2D DCT and transform domain residual DMVC with 3D DCT) together with pixel domain DMVC. In these simulations, it is proved that the transform domain DMVC can code the video content from multiple cameras more efficiently than the pixel domain DMVC. The significant improvement in R-D performance can be achieved; meanwhile the encoder still keeps low complexity which is similar as that of the conventional intra-frame coding.

6.1.3 Error Resilient Schemes for Wavelet Video Codec using WZ Coding

The increasing multimedia applications require the video content to be delivered over error prone channel. However, most of the video coding standards are suitable to storage but very sensitive to the error prone due to the utilisation of variable length coding and predictive coding etc. One usual way is to apply FEC protection to the video stream. However, this operation inevitably causes huge bandwidth consumption. In order to obtain the trade-off between error resilience and bandwidth consumption, an efficient error resilient scheme called WZER based on WZ coding is proposed in Chapter 5. The WZER algorithm is designed for the wavelet video codec and specially emphasises on the protection of ROI area in the frame. By using WZ coding principle to generate a small size of WZ stream which specially used to protect the ROI area, the proposed scheme achieves better balance between error resilience and bandwidth consumption.

The WZER uses an automatic ROI detection method to define the ROI area in the frame automatically. Then with the ROI mask generation method, the ROI related coefficients can be located with ROI mask and encoded by WZ coding. With special protection of the ROI area, the reconstruction of the video at the decoder has higher quality output in ROI area. The proposed WZER scheme has been tested with types of error-prone channel: packet erasure channel (stimulate internet transmission) and AWGN channel (stimulate the wireless transmission). The

simulation results verified the performance of WZER. Several advantages of WZER over the traditional FEC scheme are shown in the aspects of subject recovery and bandwidth consumption.

6.2 Future Works

6.2.1 Slice Structure

It is still possible to improve the WZ video coding performance with slice structure. The current slice structure presented in this thesis is not an intelligent algorithm in that the slice division is based on the restriction of the Slepian-Wolf codec used. In fact, there can be more efficient ways to divide the WZ frame into slices. It is possible to divide the slice according to the residual distribution between two adjacent WZ frames. This operation can be considered as a simple content segmentation method, with which the part of the frame with residual lower than certain threshold can be grouped into a slice. For example, with this operation, the background part will be possibly grouped into one slice. This slice can be further divided according to the restriction of the Slepian-Wolf codec. Those parts of the frame with residual higher than the predefined threshold can be grouped into one or more slices, which can be intra coded or WZ coded based on the range of the residual. This slice algorithm needs to be verified with different types of videos, e.g. fast motion or slow motion, high correlation or low correlation in both spatial and temporal directions.

6.2.2 Rate Control in WZ Video Coding

Although there are some attempts to perform rate control at the encoder or decoder for WZ coding, there are still no best solutions to solve the rate control problem. To perform rate control at the encoder is very hard. The encoder has no knowledge of the side information which results in hardly realizing rate control. However, if the side information was available at the encoder by using simple interpolation, this could cause the increase in the complexity of the system. Moreover, it is impossible to get side information available at the encoder for some applications, e.g. multiview video coding. To perform the rate control at the decoder, it is not easy to realise such algorithm since the decoder also has no information about the WZ frame transmitted. Although rate control based on the side information can be realised, the information has to be feedback to the encoder, which means the feedback channel has to be used. The feedback channel has very limited applications, which prevents the applications of WZ video coding into practical use seriously. The best solutions should be the rate control is performed at the encoder without feedback channel.

A possible research towards rate control is to design a rate control scheme at the encoder

without feedback channel. The rate control work presented in this thesis is to optimise the feedback process which can be improved with implementing the algorithm at the encoder. To perform rate control at the encoder, the side information is necessary. Although the side information may be impossible at the encoder as analyzed before, we can use the previous WZ frame as side information. In most WZ coding solution with $\text{GOP}=2$ which means there is only one WZ frame between two key frames, the side information generated by two key frames will have high correlation with the previous WZ frame. Therefore, utilizing previous WZ frame as side information, the rate control is possible to be achieved. However, this algorithm has to be verified with different video types, or different GOP lengths.

6.2.3 Error Resilience with WZ Coding

The current WZER scheme mainly works for I frame in the GOP. However, with worse channel condition, the errors happened in MV or residual in the main systematic stream can degrade the output quality seriously. The current WZER scheme cannot cover the MV or residual in systematic stream. In the future research, WZER should be explored to protect the MV and residual related to the ROI area. With this operation, the reconstruction in ROI area can achieve higher quality.

However, such operation in WZER results in following questions: Firstly, an algorithm has to be devised to track the ROI of MV and residual. Secondly, the MV and residual are produced for each P or B frame in the GOP. To protect all these MV and residual, how is the bandwidth consumption comparing to the FEC. Will WZER still show the advantages presented in this thesis? If the bandwidth consumption is too high, is it possible to compress these data since the residual and MV all are relatively small values? Such WZER scheme has to be verified again with worse channel, e.g. packet erasure channel with high packet loss rate and Rayleigh fading channel.

Reference:

- [1] D. Slepian and J. K. Wolf, "Noiseless Coding of Correlated Information Sources," *IEEE Trans. Information Theory*, vol. IT-19, pp. 471-480, 1973.
- [2] A. D. Wyner and J. Ziv, "Rate-Distortion function for source coding with side information at the decoder," *IEEE Trans. Information Theory*, vol. IT-22, pp. 1-10, 1976.
- [3] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *IEEE Proc. Special Issue on Video Coding and Delivery*, pp. 71-83, 2005.
- [4] A. D. Wyner, "Recent Results in the Shannon Theory," *IEEE Trans. Information Theory*, vol. IT-20, pp. 2-10, 1974.
- [5] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUSS): design and construction," *Data Compression Conference Proceedings*, pp. 158-167, 1999.
- [6] X. Wang and M. T. Orchard, "Design of trellis codes for source coding with side information at the decoder," *Data Compression Conference Proceedings*, pp. 361-370, 2001.
- [7] J. Bajcsy and P. Mitran, "Coding for the Slepian-Wolf problem with turbo codes," *IEEE Global Telecommunications Conf.*, pp. 1400-1404, 2001.
- [8] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Communications Letters*, vol. 5, pp. 417-419, 2001.
- [9] A. Aaron and G. B., "Compression with side information using turbo codes," *Data Compression Conference Proceedings*, pp. 252-261, 2002.
- [10] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Data Compression of Correlated nonbinary sources using punctured codes," *Data Compression Conference Proceedings*, pp. 242-251, 2002.
- [11] Y. Zhao and J. Garcia-Frias, "Joint estimation and compression of correlated nonbinary sources using punctured turbo codes," *IEEE Trans. Communications*, vol. 53, pp. 385-390, 2005.
- [12] P. Mitran and J. Bajcsy, "Coding for the Wyner-Ziv problem with turbo-like codes," *IEEE Int. Sym. Information Theory* p. 91, 2002.
- [13] J. Garcia-Frias and Y. Zhao, "Compression of binary memoryless sources using punctured turbo codes," *IEEE Communications Letters*, vol. 6, pp. 394-396, 2002.
- [14] G. C. Zhu and F. Alajaji, "Turbo codes for nonuniform memoryless sources over noisy channels," *IEEE Communications Letters*, vol. 6, pp. 64-66, 2002.
- [15] J. Garcia-Frias, "Joint source-channel decoding of correlated sources over noisy channels," *Data Compression Conference Proceedings*, pp. 283-292, 2001.
- [16] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, pp. 440-442, 2002.
- [17] T. P. Coleman, A. H. Lee, M. Medard, and M. Effros, "On some new approaches to practical Slepian-Wolf compression inspired by channel coding," *Data Compression Conference Proceedings*, pp. 282-291, 2004.
- [18] C. F. Lan, A. D. Liveris, K. Narayanan, Z. Xiong, and C. Georghiades, "Slepian-wolf coding of multiple M-ary sources using LDPC codes," *Data Compression Conference Proceedings*, p. 549, 2004.
- [19] N. Gehrig and P. L. Dragotti, "Symmetric and A-symmetric Slepian-Wolf codes with systematic and non-systematic linear," *IEEE Communications Letters*, vol. 9, pp. 61-63, 2005.
- [20] Z. Liu, S. Cheng, A. D. Liveris, and Z. Xiong, "Slepian-Wolf coded nested quantization (SWC-NQ) for Wyner-Ziv coding: Performance analysis and code design," *Data Compression Conference Proceedings*, pp. 322-331, 2004.
- [21] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes," *Asilomar Conf. Signals, Systems and Computers*, pp. 825-829, 2003.
- [22] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine*, vol. 21, pp. 80-94, 2004.
- [23] M. Fleming and M. Effros, "Network vector quantization," *Data Compression Conference*

- Proceedings*, pp. 13-22, 2001.
- [24] D. Muresan and M. Effros, "Quantization as histogram segmentation: Optimal scalar quantizer design in network systems," *IEEE Trans. Information Theory*, vol. 54, pp. 344-366, 2008.
- [25] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," *Asilomar Conf. on Signals, Systems and Computers*, pp. 240-244, 2002.
- [26] A. Aaron, E. Setton, and B. Girod, "Towards practical Wyner-Ziv coding of video," *IEEE Int. Conf. Image Processing*, pp. 869-872, 2003.
- [27] D. N. Rowitch and L. B. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo (RCPT) codes," *IEEE Trans. Communications*, vol. 48, pp. 948-959, 2000.
- [28] C. Brites, J. Ascenso, and F. Pereira, "Improving transform domain Wyner-Ziv video coding performance," *IEEE Int. Conf. Acoustics, Speech and Signal Processing 2006*.
- [29] B. Vucetic and J. Yuan, *Turbo Codes Principles and Applications*. USA: Kluwer Academic Publisher, 2000.
- [30] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near SHANNON limit error-correcting coding and encoding: Turbo-codes (1)," *IEEE Int. Conf. Communications*, pp. 1064-1070, 1993.
- [31] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform-domain wyner-ziv codec for video," *Int. Society for Optical Engineering*, pp. 520-528, 2004.
- [32] A. Aaron, D. Varodayan, and B. Girod, "Wyner-Ziv residual coding of video," *The 25th Proc. Picture Coding Symposium 2006*, 2006.
- [33] J. Ascenso, C. Brites, and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," *The 5th EURASIP Conf. Speech and Image Processing, Multimedia Communications and Services*, 2005.
- [34] Y. Vatis, S. Klomp, and J. Ostermann, "Inverse bit plane decoding order for turbo code based distributed video coding," *Int. Conf. Image Processing*, 2006.
- [35] L. Alparone, M. Barni, F. Bartolini, and V. Cappellini, "Adaptively Weighted Vector Median Filters for Motion-Fields Smoothing," *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 2267-2270, Georgia, USA, 1996.
- [36] X. Artigas, E. Angeli, and L. Torres, "Side information generation for multiview distributed video coding using a fusion approach," *The 7th Nordic Signal Processing Sym.*, pp. 250-253, 2007.
- [37] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," *ERL Technical Report*, 2003.
- [38] R. Puri and K. Ramchandran, "PRISM: A "reversed" multimedia coding paradigm," *IEEE Int. Conf. Image Processing*, pp. 617-620, 2003.
- [39] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," *Allerton Conf. Communication, Control and Computing*, 2002.
- [40] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: A video coding paradigm with motion estimation at the decoder," *IEEE Trans. Image Processing*, vol. 16, pp. 2436-2448, 2007.
- [41] R. Puri and K. Ramchandran, "Prism: An uplink-friendly multimedia coding paradigm," *IEEE Int. Conf. Acoustics, Speech and Signal Processing* pp. 856-859, 2003.
- [42] A. Wang, Y. Zhao, and H. Wang, "LVQ based distributed video coding with LDPC in pixel domain," *Lecture Notes in Computer Science* pp. 1248-1252, 2006.
- [43] J. Li, Y. Zhao, and A. Wang, "Wavelet video coding combining pyramidal lattice vector quantization with wyner-ziv coding," *Int. Conf. Signal Processing Proc. ICSP*, 2007.
- [44] F. Sheng, L. Xu-Jian, and Z. Li-Wei, "A lloyd-max-based non-uniform quantization scheme for Distributed Video Coding," *The 8th ACIS Int. Conf. Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, pp. 848-853, 2007.
- [45] W. Bo, G. Xun, Z. Debin, G. Wen, and W. Feng, "An optimal non-uniform scalar quantizer for distributed video coding," *IEEE Int. Conf. Multimedia and Expo*, pp. 165-168, 2006.
- [46] Q. Xu and Z. Xion, "Layered Wyner-Ziv video coding," *IEEE Trans. Image Processing*, vol. 15, pp. 3791-3803, 2006.
- [47] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445-3462, 1993.
- [48] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, pp. 243-250, 1996.

- [49] Y. Feng, D. Guiguang, D. Qionghai, and Y. Yaguang, "Wavelet based scalable Wyner-Ziv video coding," *The 25th Proc. Picture Coding Symposium 2006*, 2006.
- [50] M. Grangetto, E. Magli, and G. Olmo, "Context-based distributed wavelet video coding," *IEEE 7th Workshop on Multimedia Signal Processing*, 2006.
- [51] B. Wu, X. Ji, D. Zhao, and W. Gao, "Wavelet based distributed video coding with spatial scalability," *IEEE Int. Sym. Circuits and Systems*, pp. 3458-3461, 2008.
- [52] X. Guo, Y. Lu, F. Wu, and W. Gao, "Distributed video coding using wavelet," *IEEE Int. Sym. Circuits and Systems*, pp. 5427-5430, 2006.
- [53] S. Li, S. Fang, and Z. Li, "Wyner-Ziv video coding for low bitrate using SPIHT algorithm," *IEEE Workshop on Signal Processing Systems*, pp. 341-345, 2007.
- [54] Y. Tonomura, T. Nakachi, and T. Fujii, "Distributed video coding using JPEG 2000 coding scheme," *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol. E90-A, pp. 581-589, 2007.
- [55] A. Wang, Y. Zhao, and J. S. Pan, "An efficient hybrid distributed video coding," *IEICE Electronics Express*, vol. 5, pp. 650-656, 2008.
- [56] G. Wu, L. Sun, and F. Huang, "Consistent-quality distributed video coding framework," in *Lecture Notes in Computer Science* vol. 4810 LNCS Hong Kong, 2007, pp. 628-637.
- [57] M. Guo, Y. Lu, F. Wu, S. Li, and W. Gao, "Scalable Wyner-Ziv video coding with adaptive bit-plane representation," *Int. Society for Optical Engineering*, 2008.
- [58] L. Liu and E. J. Delp, "Wyner-Ziv video coding using LDPC codes," *The 7th Nordic Signal Processing Symposium*, pp. 258-261, Reykjavik, 2007.
- [59] R. P. Westerlaken, S. Bordiert, R. K. Gunnewiek, and R. L. Lagendijk, "Analyzing symbol and bit plane-based LDPC in distributed video coding," *Int. Conf. Image Processing*, 2006.
- [60] Q. Xu, V. Stankovic, and Z. Xiong, "Wyner-Ziv video compression and fountain codes for receiver-driven layered multicast," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, pp. 901-906, 2007.
- [61] C. Brites, J. Ascenso, and F. Pereira, "Modeling correlation noise statistics at decoder for pixel based Wyner-Ziv video coding," *The 25th Proc. Picture Coding Symposium 2006*, 2006.
- [62] C. Brites, J. Ascenso, and F. Pereira, "Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding," *IEEE Int. Conf. on Image Processing*, 2006.
- [63] M. Dalai, R. Leonardi, and F. Pereira, "Improving turbo codec integration in pixel-domain distributed video coding," *IEEE Int. Conf. Acoustics, Speech and Signal Processing* 2006.
- [64] L. B. Qing, X. H. He, and R. Lv, "Modeling non-stationary correlation noise statistics for Wyner-Ziv video coding," *Int. Conf. Wavelet Analysis and Pattern Recognition*, pp. 316-320, 2008.
- [65] R. P. Westerlaken, R. K. Gunnewiek, and R. L. Lagendijk, "The role of the virtual channel in distributed source coding of video," *Int. Conf. Image Processing*, pp. 581-584, 2005.
- [66] M. Guo, Y. Lu, F. Wu, S. Li, and W. Gao, "Distributed video coding with spatial correlation exploited only at the decoder," *IEEE Int. Sym. Circuits and Systems*, pp. 41-44, 2007.
- [67] S. Fang, Z. Li, and L. W. Zhang, "Distributed video codec modelling correlation noise in wavelet coarsest subband," *Electronics Letters*, vol. 43, pp. 1266-1267, 2007.
- [68] X. Artigas and L. Torres, "Iterative generation of motion-compensated side information for distributed video coding," *Int. Conf. Image Processing*, pp. 833-836, 2005.
- [69] J. Ascenso, C. Brites, and F. Pereira, "Motion compensated refinement for low complexity pixel based distributed video coding," *IEEE Int. Conf. Advanced Video and Signal Based Surveillance*, pp. 593-598, 2005.
- [70] A. Aaron, S. Rane, and B. Girod, "Wyner-Ziv video coding with hash-based motion compensation at the receiver," *Int. Conf. Image Processing*, pp. 3097-3100, 2004.
- [71] J. Ascenso and F. Pereira, "Adaptive hash-based side information exploitation for efficient Wyner-Ziv video coding," *Int. Conf. Image Processing*, 2006.
- [72] L. Natario, C. Brites, J. Ascenso, and F. Pereira, "Extrapolating side information for low-delay pixel-domain distributed video coding," *Lecture Notes in Computer Science* pp. 16-21, 2006.
- [73] A. B. B. Adikari, W. A. C. Fernando, H. K. Arachchi, and W. A. R. J. Weerakkody, "Multiple side information streams for distributed video coding," *Electronics Letters*, vol. 42, pp. 1447-1449, 2006.
- [74] W. J. Chien, L. J. Karam, and G. P. Abousleman, "Distributed video coding with lossy side information," *IEEE Int. Conf. Acoustics, Speech and Signal Processing* 2006.
- [75] W. J. Chien, L. J. Karam, and G. P. Abousleman, "Distributed video coding with 3-D recursive

- search block matching," *IEEE Int. Sym. Circuits and Systems*, pp. 5415-5418, 2006.
- [76] D. Kubasov and C. Guillemot, "Mesh-Based Motion-Compensated Interpolation for Side Information Extraction in Distributed Video Coding," *IEEE Int. Conf. Image Processing*, pp. 261-264, 2006.
- [77] B. Ko, H. J. Shim, and B. Jeon, "Wyner-Ziv video coding with side matching for improved side information," in *Lecture Notes in Computer Science*. vol. 4872 LNCS Santiago, 2007, pp. 816-825.
- [78] M. Maitre, C. Guillemot, and L. Morin, "3-D model-based frame interpolation for distributed video coding of static scenes," *IEEE Trans. Image Processing*, vol. 16, pp. 1246-1257, 2007.
- [79] S. Argyropoulos, N. Thomos, N. V. Boulgouris, and M. G. Strintzis, "Adaptive frame interpolation for Wyner-Ziv video coding," *IEEE 9th Int. Workshop on Multimedia Signal Processing*, pp. 159-162, 2007.
- [80] Y. Feng, S. Xiao, C. Wu, and R. Song, "A refined algorithm for the side information in DVC," *The 4th Int. Conf. Intelligent Information Hiding and Multimedia Signal Processing*, pp. 675-678, 2008.
- [81] X. Zhang and J. Zhangs, "Side information generation for distributed video coding based on optimal filtering," *Int. Society for Optical Engineering*, 2008.
- [82] B. T. Choi, S. H. Lee, and S. J. Ko, "New frame rate up-conversion using bi-directional motion estimation," *IEEE Trans. Consumer Electronics*, vol. 46, pp. 603-609, 2000.
- [83] S. H. Lee, O. Kwon, and R. H. Park, "Weighted-Adaptive Motion-Compensated Frame Rate Up-Conversion," *IEEE Trans. Consumer Electronics*, vol. 49, pp. 485-492, 2003.
- [84] P. Haavisto, J. Juhola, and Y. Neuvo, "Fractional frame rate up-conversion using weighted median filters," *IEEE Trans. Consumer Electronics*, vol. 35, pp. 272-278, 1989.
- [85] H. Sasai, S. Kondo, and S. Kadono, "Frame-rate up-conversion using reliable analysis of transmitted motion information," *IEEE Int. Conf. Acoustics, Speech and Signal Processing* 2004.
- [86] T. Ha, S. Lee, and J. Kim, "Motion compensated frame rate conversion by overlapped block-based motion estimation algorithm," *IEEE Int. Sym. Consumer Electronics*, pp. 345-350, 2004.
- [87] K. Kawaguchi and S. K. Mitra, "Frame rate up-conversion considering multiple motion," *IEEE Int. Conf. Image Processing*, pp. 727-730, 1997.
- [88] S. Liu, J. Kim, and C. C. J. Kuo, "Non-linear motion-compensated interpolation for low bit rate video," *Int. Society for Optical Engineering*, pp. 203-213, 2000.
- [89] G. Dane, K. El-Maleh, and Y. C. Lee, "Encoder-assisted adaptive video frame interpolation," *IEEE Int. Conf. Acoustics, Speech and Signal Processing* 2005.
- [90] A. Chang, O. C. Au, and Y. M. Yeung, "A novel approach to fast multi-frame selection for H.264 video coding," *IEEE Int. Conf. Acoustics, Speech and Signal Processing* pp. 413-416, 2003.
- [91] M. J. Chen, Y. Y. Chiang, H. J. Li, and M. C. Chi, "Efficient multi-frame motion estimation algorithms for MPEG-4 AVC/JVT/H.264," *IEEE Int. Sym. Circuits and Systems*, vol. 3, 2004.
- [92] Y. Su and M. T. Sun, "Fast multiple reference frame motion estimation for H.264/AVC," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 16, pp. 447-452, 2006.
- [93] L. Wei, Y. Zhao, and A. Wang, "Improved Side-Information in Distributed Video Coding," *Int. Conf. Innovative Computing, Information and Control*, vol. 2, pp. 209-212, 2006.
- [94] C. Brites and F. Pereira, "Encoder Rate Control for Transform Domain Wyner-Ziv Video Coding," *IEEE Int. Conf. Image Processing*, pp. II - 5-II - 8, 2007.
- [95] M. Morbee, J. Prades-Nebot, A. Pizurica, and W. Philips, "Rate allocation algorithm for pixel-domain distributed video coding without feedback channel," *IEEE Int. Conf. Acoustics, Speech and Signal Processing* Honolulu, HI, 2007.
- [96] M. Morbee, J. Prades-Nebot, A. Roca, A. Pizurica, and W. Philips, "Improved pixel-based rate allocation for pixel-domain distributed video coders without feedback channel," in *Lecture Notes in Computer Science*. vol. 4678 LNCS Delft, 2007, pp. 663-674.
- [97] A. Roca, M. Morbe?e, J. Prades-Nebot, and E. J. Delp, "Rate control algorithm for pixel-domain Wyner-Ziv video coding," *Int. Society for Optical Engineering*, 2008.
- [98] D. Kubasov, K. Lajnef, and C. Guillemot, "A Hybrid Encoder/Decoder Rate Control for Wyner-Ziv Video Coding with a Feedback Channel," *IEEE 9th Workshop on Multimedia Signal Processing*, pp. 251-254, 2007.
- [99] F. Dufaux, M. Ouaret, and T. Ebrahimi, "Recent advances in multi-view distributed video coding," *Int. Society for Optical Engineering*, 2007.
- [100] X. Guo, Y. Lu, F. Wu, W. Gao, and S. Li, "Distributed multi-view video coding," *Int. Society for Optical Engineering*, 2006.

- [101] X. Guo, Y. Lu, F. Wu, D. Zhao, and W. Gao, "Wyner-Ziv-based multiview video coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18, pp. 713-724, 2008.
- [102] L. W. Kang and C. S. Lu, "Multi-view distributed video coding with low-complexity inter-sensor communication over wireless video sensor networks," *Int. Conf. Image Processing*, pp. III13-III16, 2006.
- [103] M. Morbee, L. Tessens, J. Prades-Nebot, A. Pizurica, and W. Philips, "A distributed coding-based extension of a mono-view to a multi-view video system," *Proc. 3DTV-CON*, 2007.
- [104] M. Morbee, L. Tessens, H. Quang Luong, J. Prades-Nebot, A. Pizurical, and W. Philips, "A distributed coding-based content-aware multi-view video system," *ACM/IEEE Int. Conf. Distributed Smart Cameras*, pp. 355-362, 2007.
- [105] T. R. Natarajan and N. Ahmed, "On interframe transform coding," *IEEE Trans. Communications*, vol. COM-25, pp. 1323-1329, 1977.
- [106] A. J. Roese, "Interframe coding of digital image using transform and hybrid transform/predictive techniques," in *University of Southern California, Los Angeles. Image Processing Inst.*, 1976.
- [107] M. Bauer and K. Sayood, "Video coding using 3 dimensional DCT and dynamic code selection," *Data Compression Conference Proceedings*, p. 451, 1995.
- [108] R. K. W. Chan and M. C. Lee, "3D-DCT quantization as a compression technique for video sequences," *Annual Int. Conf. on Virtual Systems and Multimedia*, pp. 188-196, 1997.
- [109] M. C. Lee, R. K. W. Chan, and D. A. Adjero, "Quantization of 3D-DCT Coefficients and Scan Order for Video Compression," *Journal of Visual Communication and Image Representation*, vol. 8, pp. 405-422, 1997.
- [110] M. Bakr and A. E. Salama, "Implementation of 3D-DCT based Video Encoder/Decoder system," *Midwest Symp. Circuits and Systems*, pp. II13-II16, 2002.
- [111] M. Servais and G. De Jager, "Video compression using the three dimensional Discrete Cosine Transform (3D-DCT)," *The South African Sym. Communications and Signal Processing*, pp. 27-32, 1997.
- [112] B. Furht, K. Gustafsson, H. Huang, and O. Marques, "An adaptive three-dimensional DCT compression based on motion analysis," *The ACM Sym. Applied Computing*, pp. 765-768, 2003.
- [113] N. Boz?inovic and J. Konrad, "Motion analysis in 3D DCT domain and its application to video coding," *Signal Processing: Image Communication*, vol. 20, pp. 510-528, 2005.
- [114] S. Boussakta and H. O. Alshibami, "Fast Algorithm for the 3-D DCT-II," *IEEE Trans. Signal Processing*, vol. 52, pp. 992-1001, 2004.
- [115] J. J. Koivusaari and J. H. Takala, "Simplified three-dimensional discrete cosine transform based video codec," *Int. Society for Optical Engineering*, pp. 11-21, 2005.
- [116] L. Li and Z. Hou, "Multiview video compression with 3D-DCT," *The 5th Int. Conf. Information and Communications Technology*, pp. 59-61, 2007.
- [117] N. P. Sgouros, S. S. Athineos, P. E. Mardaki, A. P. Sarantidou, M. S. Sangriotis, P. G. Papageorgas, and N. G. Theofanous, "Use of an adaptive 3D-DCT scheme for coding multiview stereo images," *IEEE Int. Sym. Signal Processing and Information Technology*, pp. 180-185, 2005.
- [118] A. Smolic, K. Mueller, N. Stefanoski, J. Osterriann, A. Gotchev, G. B. Akar, G. Triantafyllidis, and A. Koz, "Coding algorithms for 3DTV - A survey(read)," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, pp. 1606-1620, 2007.
- [119] X. Guo, Y. Lu, F. Wu, W. Gao, and S. Li, "Free viewpoint switching in multi-view video streaming using Wyner-Ziv video coding," *Int. Society for Optical Engineering*, 2006.
- [120] C. Khirallah, L. Stankovic, and V. Stankovic, "Low complexity multi-view video streaming over multipath fading channels," *The 3rd Int. Conf. Information and Communication Technologies*, 2008.
- [121] Y. Li, X. Ji, D. Zhao, and W. Gao, "Region-based fusion strategy for side information generation in DMVC," *Int. Society for Optical Engineering*, 2008.
- [122] C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardi, and J. Ostermann, "Distributed monoview and multiview video coding," *IEEE Signal Processing Magazine*, vol. 24, pp. 67-76, 2007.
- [123] M. B. Badem, H. Kodikara Arachchi, S. T. Worrall, and A. M. Kondoz, "Transform Domain Residual Coding Technique for Distributed Video Coding," in *Picture Coding Sym.*, 2007.
- [124] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Processing Magazine*, vol. 17, pp. 61-82, 2000.
- [125] Y. Chen and W. A. Pearlman, "Three-dimensional subband coding of video using the zero-tree

- method," *Int. Society for Optical Engineering*, vol. 2727, pp. 1302-1312, 1996.
- [126] B.-J. Kim and W. A. Pearlman, "Embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," *Data Compression Conference Proceedings*, pp. 251-260, 1997.
- [127] C. D. Creusere, "Robust image coding using the embedded zerotree wavelet algorithm," *Proc. Data Compression Conference*, p. 432, 1996.
- [128] W.-t. Tan and A. Zakhor, "Resilient compression of video for transmission over the Internet," *Asilomar Conf. on Signals, Systems and Computers*, vol. 1, pp. 243-247, 1998.
- [129] A. A. Alatan, M. Zhao, and A. N. Akansu, "Unequal error protection of SPIHT encoded image bit streams," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 814-818, 2000.
- [130] P. Greg Sherwood and K. Zeger, "Progressive image coding for noisy channels," *IEEE Signal Processing Letters*, vol. 4, pp. 189-191, 1997.
- [131] H. Man, R. L. De Queiroz, and M. J. T. Smith, "Three-dimensional subband coding techniques for wireless video communications," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12, pp. 386-397, 2002.
- [132] Z. Xiong, B.-J. Kim, and W. A. Pearlman, "Progressive video coding for noisy channels," in *IEEE Int. Conf. Image Processing*, vol. 1, 1998, pp. 334-337.
- [133] S. Cho and W. A. Pearlman, "A full-featured, error-resilient, scalable wavelet video codec based on the set partitioning in hierarchical trees (SPIHT) algorithm," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12, pp. 157-171, 2002.
- [134] S. Cho and W. A. Pearlman, "Error resilient video coding with improved 3-D SPIHT and error concealment," *Int. Society for Optical Engineering*, pp. 125-136, 2003.
- [135] S. Cho and W. A. Pearlman, "Multilayered protection of embedded video bitstreams over binary symmetric and packet erasure channels," *Journal of Visual Communication and Image Representation*, vol. 16, pp. 359-378, 2005.
- [136] M. Tun, K. K. Loo, and J. Cosmas, "Error-resilient performance of dirac video codec over packet-erasure channel," *IEEE Trans. Broadcasting*, vol. 53, pp. 649-659, 2007.
- [137] "Dirac Software," Available at <http://diracyvideo.org/>.
- [138] A. Aaron, S. Rane, D. Rebollo-Monedero, and B. Girod, "Systematic lossy forward error protection for video waveforms," *IEEE Int. Conf. Image Processing*, pp. 609-612, 2003.
- [139] A. Aaron, S. Rane, R. Zhang, and B. Girod, "Wyner-Ziv coding for video: Applications to compression and error resilience," *IEEE Int. Conf. Data Compression* pp. 93-102, 2003.
- [140] S. Rane, A. Aaron, and B. Girod, "Systematic lossy forward error protection for error-resilient digital video broadcasting - A Wyner-Ziv coding approach," *Int. Conf. Image Processing*, pp. 3101-3104, 2004.
- [141] S. Rane and B. Girod, "Systematic lossy error protection versus layered coding with unequal error protection," *Int. Society for Optical Engineering*, pp. 663-671, 2005.
- [142] S. Rane and B. Girod, "Systematic lossy error protection of video based on H.264/AVC redundant slices," *Int. Society for Optical Engineering*, 2006.
- [143] P. Baccichet, S. Rane, and B. Girod, "Systematic lossy error protection based on H.264/AVC redundant slices and flexible macroblock ordering," *Journal of Zhejiang University: Science*, vol. 7, pp. 900-909, 2006.
- [144] S. Rane, A. Aaron, and B. Girod, "Error-resilient video transmission using multiple embedded Wyner-Ziv descriptions," *Int. Conf. Image Processing*, pp. 666-669, 2005.
- [145] O. Crave, C. Guillemot, B. Pesquet-Popescu, and C. Tillier, "Distributed temporal multiple description coding for robust video transmission," *Eurasip Journal on Wireless Communications and Networking*, vol. 2008, 2008.
- [146] A. Sehgal, A. Jagmohan, and N. Ahuja, "Wyner-Ziv Coding of Video: An Error-Resilient Compression Framework," *IEEE Trans. Multimedia*, vol. 6, pp. 249-258, 2004.
- [147] M. Wu, D. Song, and C. W. Chen, "Multiple description distributed image coding with side information for mobile wireless transmission," *Int. Society for Optical Engineering*, pp. 572-582, 2005.
- [148] Q. Xu, V. Stankovic, and Z. Xiong, "Layered Wyner-Ziv video coding for transmission over unreliable channels," *Signal Processing*, vol. 86, pp. 3212-3225, 2006.
- [149] Z. Xue, K. K. Loo, and J. Cosmas, "Bandwidth efficient error resilience scheme for wavelet based video transmission," *IEEE Int. Sym. Broadband Multimedia Systems and Broadcasting 2008*, 2008.

- [150] D. Nister and C. Christopoulos, "Lossless region of interest coding," *Signal Processing*, vol. 78, pp. 1-17, 1999.
- [151] S. Y. Chien, Y. W. Huang, B. Y. Hsieh, S. Y. Ma, and L. G. Chen, "Fast video segmentation algorithm with shadow cancellation, global motion compensation, and adaptive threshold techniques," *IEEE Trans. Multimedia*, vol. 6, pp. 732-748, 2004.
- [152] M. N. Gurcan, Yardimci, Yasemin, Cetin, A. Enis, "Influence function based Gaussianity tests for detection of microcalcifications in mammogram images," in *IEEE Int. Conf. Image Processing*. vol. 3, 1999, pp. 407-411.
- [153] C. Christopoulos, J. Askelo?f, and M. Larsson, "Efficient methods for encoding regions of interest in the upcoming JPEG2000 still image coding standard," *IEEE Signal Processing Letters*, vol. 7, pp. 247-249, 2000.
- [154] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM Journal on Mathematical Analysis*, vol. 29, pp. 511-546, 1998.
- [155] P. Getreuer, "Filter Coeffcients to Popular Wavelets,"
<http://www.mathworks.com/matlabcentral/fileexchange/5502>, 2006.
- [156] C. P. Ho and C. J. Tsai, "Content-adaptive packetization and streaming of wavelet video over IP networks," *Eurasip Journal on Image and Video Processing*, vol. 2007, 2007.
- [157] J. Kim, R. M. Mersereau, and Y. Altunbasak, "Error-resilient image and video transmission over the Internet using unequal error protection," *IEEE Trans. Image Processing*, vol. 12, pp. 121-131, 2003.
- [158] M. Wu, S. S. Karande, and H. Radha, "Network-embedded FEC for optimum throughput of multicast packet video," *Signal Processing: Image Communication*, vol. 20, pp. 728-742, 2005.
- [159] "LDPC-Staircase code," Available at: <http://planete-bcast.inrialpes.fr/>.
- [160] B. M. J. Leiner, "LDPC Codes – a brief Tutorial," April 8 2005.
- [161] R. G. Gallager, "Low Density Parity-Check Codes," *MIT Press*, 1963.
- [162] J. Pearl, "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference," *MorganKaufmann Publishers*, 1988.