STATUS OF THESIS

Title of thesis	REGION-BASED ADAPTIVE DISTRIBUTED VIDEO CODING CODEC

I, ABDELRAHMAN ELAMIN ABDELRAHMAN ELAMIN

hereby allow my thesis to be placed at the Information Resource Center (IRC) of Universiti Teknologi PETRONAS (UTP) with the following conditions:

1. The thesis becomes the property of UTP
2. The IRC of UTP may make copies of the thesis for academic purposes only.
3. This thesis is classified as

[ ]	Confidential
[✓]	Non-confidential

If this thesis is confidential, please state the reason:

____

The content of the thesis will remain confidential for ________ years.

Remarks on disclosure:

____

Endorsed by

Signature of Author

Signature of Supervisor
Assoc.Prof. Dr. Varun Jeoti

Permanent address:
Elfayhaa, Khartoum, North
Khartoum Sudan

Date: 21/12/2010

Date: 19th Dec 2010

UNIVERSITI TEKNOLOGI PETRONAS

Region-Based Adaptive

Distributed Video Coding Codec

By

ABDELRAHMAN ELAMIN ABDELRAHMAN ELAMIN

The undersigned certify that they have read, and recommend to the Postgraduate Studies Programme for acceptance, for the fulfillment of the requirements for the DEGREE OF DOCTOR OF PHILOSOPHY IN ELECTRICAL AND ELECTRONIC ENGINEERING.

Signature:

Main Supervisor: Assoc. Prof. Dr. Varun Jeoti

Signature:

Co-Supervisor: Dr. Samir Ibrahim Belhouari

Signature:

Head of Department: Dr. Nor Hisham Bin Hamid

Dr Nor Hisham Hamid
Head of Department
Electrical & Electronic Engineering
Universiti Teknologi PETRONAS

Date: 21/12/2010

REGION-BASED ADAPTIVE DISTRIBUTED VIDEO CODING CODEC

By

ABDELRAHMAN ELAMIN ABDELRAHMAN ELAMIN

A Thesis

Submitted to the Postgraduate Studies Programme

as a Requirement for the Degree of

DOCTOR OF PHILOSOPHY

ELECTRICAL AND ELECTRONIC ENGINEERING PROGRAMME

UNIVERSITI TEKNOLOGI PETRONAS

BANDAR SERI ISKANDAR

PERAK

December, 2010

DECLARATION OF THESIS

Title of thesis	REGION-BASED ADAPTIVE DISTRIBUTED VIDEO CODING CODEC

I, ABDELRAHMAN ELAMIN ABDELRAHMAN ELAMIN,

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurren
tly submitted for any other degree at UTP or other institutions.

Witnessed by

Signature of Author
Permanent address:
Elfayhaa, Kahrtoum North
Khartoum Sudan
Date: 21/12/2010

Signature of Supervisor
Assoc. Prof. Dr. Varun Jeoti

ACKNOWLEDGEMENTS

First I would like to thank God the Almighty, for without his consent, it would be impossible to achieve what had been done in this work. And also I would like to thank my supervisors Assoc. Prof. Dr. Varun Jeoti and Dr Samir Belhaouari brahim who push me hard to overcome the complications of this investigation with all of their knowledge.

Also the author would like to thank the management and authorities of UNIVERSITI TEKNOLOGI PETRONAS for their constant support and encouragement for the accomplishment of this research study.

DEDICATION

To my family

ABSTRACT

The recently developed Distributed Video Coding (DVC) is typically suitable for the applications where the conventional video coding is not feasible because of its inherent high-complexity encoding. Examples include video surveillance using wireless/wired video sensor network and applications using mobile cameras etc. With DVC, the complexity is shifted from the encoder to the decoder.

The practical application of DVC is referred to as Wyner-Ziv video coding (WZ) where an estimate of the original frame called "side information" is generated using motion compensation at the decoder. The compression is achieved by sending only that extra information that is needed to correct this estimation. An error-correcting code is used with the assumption that the estimate is a noisy version of the original frame and the rate needed is certain amount of the parity bits. The side information is assumed to have become available at the decoder through a virtual channel. Due to the limitation of compensation method, the predicted frame, or the side information, is expected to have varying degrees of success. These limitations stem from location-specific non-stationary estimation noise. In order to avoid these, the conventional video coders, like MPEG, make use of frame partitioning to allocate optimum coder for each partition and hence achieve better rate-distortion performance. The same, however, has not been used in DVC as it increases the encoder complexity.

This work proposes partitioning the considered frame into many coding units (region) where each unit is encoded differently. This partitioning is, however, done at the decoder while generating the side-information and the region map is sent over to encoder at very little rate penalty. The partitioning allows allocation of appropriate DVC coding parameters (virtual channel, rate, and quantizer) to each region. The resulting regions map is compressed by employing quadtree algorithm and communicated to the encoder via the feedback channel. The rate control in DVC is performed by channel coding techniques (turbo codes, LDPC, etc.). The performance of the channel code depends heavily on the accuracy of virtual channel model that

models estimation error for each region. In this work, a turbo code has been used and an adaptive WZ DVC is designed both in transform domain and in pixel domain. The transform domain WZ video coding (TDWZ) has distinct superior performance as compared to the normal Pixel Domain Wyner-Ziv (PDWZ), since it exploits the spatial redundancy during the encoding. The performance evaluations show that the proposed system is superior to the existing distributed video coding solutions. Although the proposed system requires extra bits representing the "regions map" to be transmitted, but still the rate gain is noticeable and it outperforms the state-of-the-art frame based DVC by 0.6-1.9 dB.

The feedback channel (FC) has the role to adapt the bit rate to the changing statistics between the side information and the frame to be encoded. In the unidirectional scenario, the encoder must perform the rate control. To correctly estimate the rate, the encoder must calculate typical side information. However, the rate cannot be exactly calculated at the encoder, instead it can only be estimated. This work also proposes a feedback-free region-based adaptive DVC solution in pixel domain based on machine learning approach to estimate the side information. Although the performance evaluations show rate-penalty but it is acceptable considering the simplicity of the proposed algorithm.

ABSTRAK

Kaedah "Distributed Video Coding" (DVC) yang diperkenalkan mutakhir ini adalah sesuai untuk aplikasi-aplikasi yang tidak dapat dilaksanakan oleh pengkodan video tradisional yang mengandungi teknik pengkodan kompleks. Contoh-contoh aplikasi yang boleh menerima manafaat dari teknik DVC termasuklah surveilans video yang menggunakan rangkaian pengesan tanpa wayar/dengan wayar, serta aplikasi yang menggunakan kamera mudah alih, dan sebagainya. Dengan DVC, kompleksiti beralih dari pengkod ke penyahkod.

Aplikasi praktikal menggunakan DVC disebut sebagai pengkodan video Wyner-Ziv (WZ); dimana anggaran rangka asli yang dikenali sebagai "maklumat sampingan" dianggarkan menggunakan gerak pampasan di penyahkod. Mampatan terhadap maklumat dilakukan dengan hanya menghantar maklumat tambahan yang difikirkan penting dalam menambahbaik penganggaran ini. Kod pembetulan ralat digunakan dengan mengandaikan bahawa anggaran tersebut adalah versi kerangka asal yang sudah dicemari hingar, dan kadar yang diperlukan adalah dipengaruhi oleh jumlah tertentu bit-bit pariti. Maklumat sampingan dianggap telah wujud di penyahkod melalui saluran maya. Disebabkan oleh kaedah pampasan yang terbatas, kerangka teranggar atau maklumat sampingan adalah dijangka menghasilkan peringkat kejayaan pelbagai. Keterbatasan ini berasal dari penganggaran hingar bukan pegun berorientasikan lokasi khusus. Dalam mengatasi keterbatasan ini, pengkod video konvensional seperti MPEG menggunakan kerangka berpatisyen untuk memperuntukkan pengkod optimum kepada setiap partisyen, dan mencapai prestasi "kadar-herotan" yang lebih baik. Namun demikian, kaedah kerangka berpatisyen ini belum lagi digunapakai dalam DVC, kerana skim tersebut boleh meningkatkan kompleksiti pengkod.

Dalam karya ini, kami mencadangkan agar kerangka berpartisyen dipecahkan kepada unit-unit (wilayah atau daerah), dimana setiap unit diberikan kod yang unik—berbeza daripada kod yang diberikan kepada unit-unit lain. Skim pecahan

partisyen ini dilakukan di penyahkod seiring dengan penjanaan maklumat sampingan. Sementara itu, peta unit (wilayah) yang dihantar ke pengkod hanya mengalami kadar penalti yang sangat sedikit. Skim partisyen ini juga membolehkan peruntukan parameter-parameter DVC (saluran maya, kadar, kuantisasi) yang bersesuaian untuk wilayah masing-masing. Peta-peta daerah yang dihasilkan di mampatkan dengan menggunakan algoritma "Quadtree" dan peta-peta tersebut dihantar kepada pengkod melalui saluran maklum balas. Kadar kawalan dalam DVC dilaksana menggunakan teknik pengkodan saluran (kod turbo, "LDPC", dll). Prestasi kod saluran sangat bergantung kepada ketepatan model saluran maya yang memodelkan penganggaran ralat untuk setiap daerah. Dalam penyelidikan ini, kod turbo telah digunakan dan DVC WZ adaptif direkabentuk dalam domain penjelmaan (transformasi) dan domain piksel. Pengkodan video berasaskan domain transformasi Wyner-Ziv ("Wyner-Ziv Transform Domain" - TDWZ) mempunyai prestasi yang unggul berbanding dengan domain piksel Wyner-Ziv ("Wyner-Ziv Pixel Domain" - PDWZ); ini adalah kerana TDWZ memanfaatkan redundansi ruang semasa proses pengkodan. Penilaian prestasi menunjukkan bahawa sistem yang dicadangkan adalah lebih unggul berbanding dengan penyelesaian-penyelesaian pengkodan video teragih semasa. Walaupun sistem yang dicadangkan memerlukan bit tambahan yang mewakili "peta-peta wilayah" untuk dihantar, tetapi kadar keuntungan skim yang dicadangkan tetap menyerlah; prestasinya melebihi teknik asal DVC sebanyak 0,6 hingga 1.9 dB.

Saluran maklum balas ("Feedback Channel" - FC) mempunyai peranan untuk menyesuaikan kelajuan bit dengan statistik yang berubah-ubah di antara maklumat sampingan dan kerangka yang akan dikodkan. Dalam senario sehala, pengkod mestilah melakukan pengawalan keatas "kadar". Pengkod mestilah menghitung maklumat sampingan lazim (tipikal) supaya kadar boleh dihitung secara tepat. Bagaimanapun, kadar pada pengkod tidak boleh dikira setepat-tepatnya; sebaliknya, kadar hanya boleh dianggarkan. Dalam karya ini, kami juga mencadangkan penyelesaian DVC adaptif tanpa maklum balas ("feedback-free") serta berasaskan wilayah ("region-based") yang menggunakan domain piksel. Seterusnya, domain piksel ini menggunakan pendekatan pembelajaran mesin untuk menganggarkan maklumat sampingan. Walaupun penilaian prestasi menunjukkan kadar penalti, teknik yang dicadangkan boleh diterima pakai memandangkan tahap kesederhanaan algoritmanya.

COPYRIGHT PAGE

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Nomenclature

X	Original source information.
Y	Correlated source to the original information "Side information".
P	Probability
$\hat{X}$	Reconstructed version of the original source.
$\hat{Y}$	Reconstructed version of the side information.
R_x	Rate of transmitting original source.
R_Y	Rate of transmitting side information.
$R_{X\|Y}$	Rate of transmitting original source given Y is available at the decoder.
R_{WZ}^*	Wyner-Ziv rate.
Z	Residual between original source and side information.
D	Distortion.
q	Quantized symbol.
$\grave{q}$	Decoded quantized symbol.
a_0, a_1	Linear estimation parameters.
X_{2i-1}	Previous frame.
X_{2i+1}	Next frame.
MV_F	Forward motion vector.

MV_B	Backward motion vector.
P_F	Motion compensated frame using forward motion vector.
P_B	Motion compensated frame using backward motion vector.
F_1^A, F_1^B	Sum images of frame indexed 1.
α	Laplcian distribution parameters.
K_t	Quantizer depth for DCT coefficient number t.
Y_{enc}	Encoder version side information.
Y_{dec}	Decoder version side information.
L	Number of regions in the frame.
X_l	Region number l at the original frame.
Y_l	Region number l at the side information.
μ_i	Mean for partition i.
τ_i	Threshold value for partition i.
Γ_{BF}	Motion compensation operator.
σ_l^2	Variance of the residual between original frame and side information.
Q_l	Quantizer depth for region l.
ΔD	Distortion reduction.
R_k	Rate of bitplane indexed by k
O	Length of bitplane
k	Index of the bit plane.
L_n	Logarithm of the A Posterior Probability.
u_n	Input bit number n to the Turbo encoder.
u_n^P	Turbo encoder output parity bit corresponding to input bit n.

u_n^S	Turbo encoder output systematic bit corresponding to input bit n.
$\breve{\alpha}_n$	Forward probability.
S_n	State n of the RSC encoder.
$\breve{\beta}_n$	Backward probability.
$L^{apriori}$	A priori information in the SISO decoders.
r_n^P	Received parity bit corresponding to bit n
r_n^S	Received systematic bit corresponding to bit n
γ	Transition between states probability.
P_e	Error probability.
Z_l^{DCT}	Residual between original and side information in DCT domain.
B_t	DCT band number t
b_k	Bit number k.
V_{DC}	DC coefficient value
W_{DC}	DC band quantization step.
V_{AC}	AC coefficient value
W_{AC}	AC band quantization step.
ΔR	Extra rate to be estimated by the encoder.
ρ	Puncturing pattern.
Ψ	Puncture period
Đ	Delay in the RSC encoder.
C_{out}	RSC encoder input.
C_{in}	RSC encoder input.
G	RSC generator matrix.

CHAPTER 1
INTRODUCTION

This chapter presents an overview of the entire thesis. Section 1.1 covers the background of the distributed video coding (DVC) and briefly introduces one of its many applications. Present day DVC implementations are presented in section 1.2. Section 1.3 states the problem that the author intends to tackle, targeting better practical DVC systems implementation. In the light of that, section 1.4 defines the research objectives undertaken in the thesis. Section 1.5 briefly presents the approach used, the scope and brief methodology of the performance evaluation and testing of the proposed technique. Section 1.6 explores the thesis explaining how each chapter is presented and what they contain.

1.1 Background

Conventional digital video coding paradigm is represented by the ITU-T and MPEG standards mainly based on a hybrid of inter-frame predictive coding that reduces temporal redundancy and block-based transform that reduces the spatial redundancy in the video stream. In this coding framework, the encoder side is tasked to exploit both the temporal and spatial redundancies present in the video sequence. This is a complex process and requires a considerable amount of resources (both power and memory). As a result, all standard based conventional video encoders have a very high computational complexity, typically five to ten times more complex than their decoders [2, 3]. This is mainly on account of exploiting the temporal correlation, a task performed by estimating motion. It is after all the job of the encoder to do all coding decisions and work in order to achieve the best performance, while the decoder remains a pure executer of the encoder "instructions". This type of architecture is well-suited for scenarios where the video is encoded once and decoded many times. Examples include, among others, broadcasting and video-on-demand, where the cost of the decoder is more critical than that of the encoder. Recently new emerging applications such as wireless video sensor networks and multimedia sensor networks, wireless PC cameras and mobile camera phones have sprung up which represents a real challenge to the traditional video coding architecture. These applications have very different requirements than those of conventional video delivery systems. For some applications, it is essential to have lower power consumption both at the encoder and decoder sides, e.g. in mobile camera phones. In other types of applications, notably when there is a high number of encoders and only one decoder, e.g. surveillance, low cost encoder devices are necessary. In order to fulfill these requirements, it is essential to have a low-power and low-complexity encoding device, possibly at the expense of a higher complexity decoder. As a result, a lower complexity encoder could be achieved by moving some of the encoder tasks to the decoder part, specially the most complex motion estimation process. This useful hint is the consequence of information-theoretic principles established in the 1970s by Slepian and Wolf for distributed lossless coding [4], and by Wyner and Ziv for lossy coding [5] with some side information made available at the decoder. A new coding scheme referred to as distributed source coding (DSC) algorithms have emerged based on these theorems. The attractive idea of distributed source coding is that, in the case of joint decoding, two correlated sources X and Y can be compressed separately without the knowledge of the other source except the knowledge of their correlation at both the encoder and the decoder. It can still attain the same compression as if the other source was known. For the specific case where Y is available at the decoder and X and Y are jointly Gaussian, Wyner [5] proved that by using channel coding at the encoder the compression can still be achieved without Y being known at the encoder, since the correlated source and dependency between the two sources are available at the decoder. In practice, the dependency between X and Y is characterized as a virtual dependency channel described by the transition probability function $P(X|Y)$. This characterization of the dependency leads to the insight that Slepian-Wolf (SW) coding scheme can be used by 'protecting' X for 'transmission' over the virtual dependency channel $P(X|Y)$. Recent researches have proposed that the concept of Wyner-Ziv (WZ) coding can be applied to video coding as well, allowing a low-encoding complexity at the cost of a more computationally demanding decoder [6-11]. In these studies, two different error correcting codes have been studied. Works in [6, 9, 10, 12] have used trellis syndrome codes to construct their PRISM video codec, whereas work in [11] has studied the ability of a turbo code based system to SW code video data. These studies introduced distributed video coding (DVC) as a new coding paradigm that allows moving certain parts of the encoder module to the decoder, thus reducing the complexity at the encoder. This means the overall video coding complexity allocation can be flexible, depending on the conditions at hand, like available power on both sides and the available bandwidth.

As mentioned earlier, Wireless Video Sensor Network (WVSN) is a fast emerging application. WVSN perform the function of the video coding in very tight constraints of limited communication bandwidth, processing capability and power supply. Typical WVSN consists of hundreds of video sensors connected to a central node - a high number of encoders and only one decoder. When a high number of encoders is needed, the low cost of the encoder is mandatory. Therefore, it is essential to have low-power consumption, thus low-complexity, possibly at the expense of a higher complexity decoder. DVC, in such a limited resources environment, proves to be an advantageous solution. The main advantage of using DVC in this scenario is that DVC enables cheap and light encoders to efficiently perform the video compression task. Another advantage is in a multi-view video context where there exists inter-view correlation among various views of the same scene. The conventional video coding paradigm, used in this scenario, requires cameras to communicate among themselves to exploit the interview correlation at the joint encoder. However, using the DVC approach, a significant amount of such communication is avoided and complexity reduced, for DVC-based encoders do not need to jointly process the different views and thus do not need inter-camera and inter-encoder communication.

1.2 Present Day DVC Implementations

Current DVC systems tackle the problem of distributed video coding at two levels of granularity - block level and frame level. DVC systems using block level granularity are called block based systems such as PRISM [6, 9, 12] whereas those using frame level granularity are called frame based systems such as those in [2, 3, 8]. The block based system is developed by the Berkeley university research group where the video frame is divided into 8-by-8 blocks. These blocks are classified into three types of blocks (SKIP, INTRA, INTER) based on the mean-squared-error (MSE) between the current block and co-located block in the previous frame, WZ encoding is only applied on the INTER blocks. A typical video compression scheme exploits both the temporal and spatial correlations. In PRISM, the temporal correlations are utilized in the SKIP blocks – the blocks that are simply replaced by the co-located block in the previous frame while INTER blocks, encoded WZ, are decoded by performing encoder-guided motion compensation. The classification process, typically performed by comparing the MSE of co-located blocks in two temporally adjacent frames, represents undesired complexity in the low–complexity encoding. In the conventional video coding, the SKIP-mode is applied to the blocks with zero motion vectors after performing motion estimation [9].

An alternative scheme based on frames has been developed at Stanford University [3, 7, 8, 13] and this scheme has become the most popular WZ video codec. The basic idea of this WZ video coding architecture is that the decoder based on some previously and conventionally transmitted frames, popularly known as key frames, creates the so-called side information (SI) that works as estimates for the other frames to be encoded. The WZ frames are encoded using a channel coding approach. Channel codes like turbo codes or Low-Density Parity-Check (LDPC) codes [14, 15] create redundant parity bits which are used at the decoder to correct the 'estimation' errors in the corresponding decoder frames. In this case, the encoding is performed assuming there is (a high) correlation between the original WZ frames to code and their associated SI frames at the decoder. The higher this correlation is, the more efficient this encoding process becomes.

1.3 Problem Statement

In the conventional video coding, the current frame, X, is encoded with respect to certain reference frame, Y. This reference frame is the previous frame and it is made available at the decoder. In WZ video coding, the reference is an estimate of the current frame at the decoder. A close estimate for the current frame is generated at the decoder by exploiting the temporal correlation in video sequences; motion estimation (ME) algorithms can be used to exploit this correlation. Most of the frame prediction algorithms assume translational and symmetric motion between the frames for the sake of simplicity. The true motion assumes different motion model such as affine, translational or rotational models. Therefore, the frame prediction is expected to have spatially varying degrees of success across the predicted frame. In both coding applications (the traditional video codec and the DVC codec) the optimal video coding solution is the one that considers the local characteristics of the video sequence. In conventional video coding, an already mature video coding paradigm, such optimal schemes have been made use of, and they are referred as region-Based video coding. The video frame is seen as a group of different units/regions each with unique characteristics.

In the Region-Based video coding, the frame is partitioned into many coding units each coding unit has its own coding parameters, so appropriate rate, quantizer and hence rate-distortion performance. Employing the Region-Based video coding scheme in the DVC is not trivial task considering the constraint of low-complexity encoder where the parameters have to be estimated as in traditional Region-Based video coding. The challenge of implementing Region-Based video coding scheme in the DVC appears in the complexity of estimating the coding parameters for each coding unit/region. Since the estimation requires the reference frame to be available at the encoder, in the DVC scenario this reference frame (side information) is unavailable at the encoder. The relevant coding parameters in the DVC application are the dependency channel model, the quantizer depth and the bitrate, and they have great impact on the performance of the DVC codec. For example the Region-Based caters for the location-specificity of the dependency channel. The noise process in the occluded region will have a large variance and follows, for instance, a uniform probability density function (PDF) model. In the non-occluded regions, however, the noise process is assumed well-behaved, and follows a Gaussian or Laplacian PDF model with a small variance [16].

The second parameter to be estimated for each coding unit/region is the quantizer depth. The Region-Based video coding allows allocating appropriate quantizer for each region. For example the successfully estimated regions with coarser quantizer can achieve same distortion reduction as the poorly estimated regions with finer quantizer. Therefore to achieve a certain operational characteristics in terms of distortion reduction, the quantizer must consider the quality of the side information. It is imperative that, in reconstructing the encoded region, the joint decoder should rely more on side information when the rate is too low. On the other hand, when the rate is high, the decoded quantized symbol of the region becomes more reliable than side information, so the decoder should put more weight on the quantized symbol.

The third coding parameter for each coding unit in the DVC is the rate, which is estimated on the fly at the decoder by making use of a rate-compatible-punctured codes (RCPT) codec as the module of the SW codec. The encoder "blindly" sends 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.

Developing Region-Based DVC codec requires performing all Region-based video coding relevant processes such as, partitioning video frame and estimating the associated coding parameters for each coding unit. The outcome of these two processes must be available at the encoder, considering the DVC constraints defining the relevant coding parameters and estimating those parameters is quite challenge if the encoding low-complexity constraint is to be adhered to. It can be concluded that developing DVC video coding that considers the local characteristics of the video sequence by incorporating Region-Based approach is a challenging research problem.

1.4 Thesis Objective

The overall objective of this work is to design a DVC solution that makes use of the local characteristics of the video and employs the Region-based video coding in the DVC system while preserving the simplicity of the video encoder. The targeted video solution is sought to be deployed in two application scenarios bi-directional scenario with the feedback channel and uni-directional scenario without feedback.

I. Bi-directional scenario: This scenario can be itemized in the following objectives

a. To obtain more accurate channel model that considers the non-stationary nature of the dependency channel for the entire frame. Since the SW codec depend heavily on the accuracy of the dependency channel model, the overall system performance can be significantly improved.
b. To keep the localized errors close to each other in the generated symbol stream by partitioning the bitplanes into partitions that have uniform noise model, so the system could have a tighter control on the rate control process.
c. To minimize the overall reconstruction distortion by applying adaptive source coding process "Quantization" based on the spatial variation of the local video characteristics.

II. Uni-directional Scenario: The objective in this scenario is the removal of the feedback channel while adhering to the same philosophy of region-based coding solution of bi-directional scenario. The absence of feedback channel requires estimating the rate per region at the encoder without adding too much extra complexity at the encoder.

1.5 Methodology

In this study, a MATLAB™ simulation of the proposed DVC scheme is developed, while the hardware implementation of this work is out of the scope of this thesis. Video sequences in the uncompressed YUV/QCIF format are used as simulations materials. These video sequences contain different numbers of frames but for the simulation only the first 101 frames are considered. The temporal resolution (number of frames per second) is fixed at 30 frames/second during this simulation for all materials. The rate-distortion performance of the proposed solution is tested only using the luminance data for these video sequences. The group of picture (GOP) is chosen to be 2 during the simulation in this group. The odd frame is conventionally encoded and the even frame is encoded by the proposed solutions. The number of regions for each WZ frame (even frame) is determined experimentally and varies between 4 to 5 regions based on the video sequence characteristics. Turbo codes simulation borrowed from the Code Modulation Library (CML) with two parallel recursive systematic encoders (RSC). The numbers of iterations run by Turbo decoder is fixed at 20 iterations. To evaluate the proposed solution RD performance the rate-distortion curves contain the rate and the PSNR values for the even frames of the given sequence. MPEG (I-P-I-P) as conventional video coding is used for comparison to evaluate the proposed system.

1.6 Thesis Organization

The thesis is composed of following:

CHAPTER 1: Starting with a brief introduction of current mature traditional video coding standards, some important applications are highlighted where most standard video coding schemes are not applicable. Then the idea of DVC is presented and the distinct advantages are listed compared to the current traditional video coding standards. In addition, aims, objectives and main contributions of the research are presented in detail. Finally, the organization of the thesis is stated with the overall description of each chapter.

CHAPTER 2: This chapter provides the theoretical background and covers the basics concept of distributed source coding (DSC) and describes existing popular DSC theorems. Because distributed source coding can be regarded as a special case of channel coding, some channel coding methods that are widely used in the distributed source coding are also presented. The chapter also gives an overview of basic distributed video coding that is developed based on the DSC principle and covers in more details the modules of the basic DVC systems. Several existing distributed video coding systems are also reviewed in this chapter. The chapter also highlights the related literature, highlights at challenges and discusses related work to cope with these issues.

CHAPTER 3: This chapter is mainly devoted to introducing the proposed solution and its feature. The chapter also presents objectives of the proposed approach. The theoretical foundation of the Region-Based Adaptive DVC codec in pixel domain and transform domain is given.

CHAPTER 4: This chapter introduces the Region-Based Adaptive DVC codec without feedback DVC solution in pixel domain. The process of setting up encoder rate control module (offline) and attaching the module to the state-of-the-art feedback based encoder is presented in detail.

CHAPTER 5: This chapter presents the test materials (video sequences) and performance evaluation parameters. The chapter also presents, in detail, the performance evaluation methodology of the proposed Region-Based Adaptive DVC codec in pixel domain and transforms domain solutions. The methodology for Region-Based Adaptive DVC codec without Feedback solution performance evaluation is also covered in this chapter.

CHAPTER 6: In this chapter, the simulation process is first conducted to evaluate the performance of the Region-Based Adaptive DVC codec solution in pixel and transform domains. The comparisons of the performance of the proposed solution with the existing frame based DVC solution and the conventional video coding MPEG is presented in this chapter. Simulation processes are also conducted to evaluate the performance of the Region-Based Adaptive DVC codec without feedback. A comprehensive discussion for the testing procedures and results is given for all experiments.

CHAPTER 7: 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

THEAORETICAL FOUNDATIONS OF DVC

One of the key problems in video signal processing is that the data rate of the raw video signal is very large. For example, one second of Common Intermediate Format (CIF) color video (352 by 288 pixels) at a frame rate of 30 frames per second (fps) which is stored with 24 bits per pixel (bpp), i.e., 8 bb for each red (R), green (G) and the blue (B) channel, requires $352 \times 288 \times 24 \times 30 = 73$ Mbits. For a higher resolution video such as 1080p (1920 by 1080 pixels, progressive scan) High-Definition (HD) video at 30 fps, the raw video data rate is as high as $1920 \times 1080 \times 24 \times 30 = 1493$ Mbits/second (Mbps). Therefore, the raw video data rate is high and it is difficult to record or transmit directly so the compression is required.

Most of digital video coding systems are built based on source coding methods. Source coding is a technique used to compress the source, by removing redundancy and in turn, reducing the bandwidth needed for the transmission through a channel. A source characteristic such as signal redundancy is exploited in the process of compression. Video coding can be categorized into single-source data compression (intraframe coding) and multiple-source data compression (interframe coding). Intraframe coding is performed by removing the spatial redundancy within the single frame. Video sequence has a significant amount of redundancy along the time axis (adjacent frames), this redundancy is removed by interframe video coding. The interframe coding has a higher compression performance than intraframe coding with the cost of the higher complexity at the encoding part. Most of the conventional video coding systems are interframe video coding systems.

A novel video coding approach, characterized by intraframe encoding and interframe decoding, called distributed video coding, is built based on the Slepian-Wolf (SW) and the Wyner-Ziv (WZ) theorems. For many applications, including portable multimedia devices and wireless video sensor networks, it is desirable to design low-complexity video encoding systems that can result in reduced power consumption, cost and size. Conventional video coding systems such as MPEG and ITU-T H.26x, can achieve high compression ratios through computationally complex motion estimation operations at the encoder. The main objective of the DVC is to reduce the computational complexity at the encoder by shifting the temporal correlation exploitation to the decoder, while attempting to achieve the same coding performance as the conventional video coding systems.

DVC is based on information-theoretic principles established in the 1970s by Slepian [4] and by Wyner and Ziv [5]. The main goal of this chapter is to exhaustively explore the literature on WZ video codec. Section 2.1 covers the theoretical foundation of the DSC principle and the two major Information Theory results, namely the Slepian-Wolf and Wyner-Ziv theorems. Section 2.2 explains how the source coding, channel coding and estimation interplay to construct a distributed video coding solution. In section 2.3, some of the architectural developments which are derived from the basic architecture are given. Section 2.4 presents a detailed review of the two DVC practical solutions in the literature. The first is a practical WZ video codec, known as "Power-efficient, Robust, hIgh-compression, Syndrome-based Multimedia coding" PRISM, which was developed at UC Berkeley University. The second is a feedback channel based DVC which was developed by Bernd Girod's group at Stanford University. The chapter is concluded by highlighting some of the issues in existing DVC approaches as reported in the literature.

2.1 Theoretical Foundation of Distributed Video Coding

Source coding can be classified as lossless source coding and lossy source coding. Compression of a signal, wherein the decompression reproduces the original signal is referred to as lossless source coding. Slepian-Wolf [4] coding is viewed as lossless source coding. The compression of a signal with the removal of redundant information is lossy source coding, the best example of which is Wyner-Ziv coding. Shannon showed that there is a limit to which a source can be compressed without introducing errors at the decoder [17]. The rate distortion theory gives a tradeoff between compression and quality in lossy source coding. Images, video and audio are often compressed using lossy source coding to achieve better compression techniques, however compressors will have a lossless mode. An application of error correcting codes for data compression is first studied by Shannon where he suggests that there is duality between source coding and channel coding. Recently the problem of source coding using channel coding is receiving increasing attention. This problem occurs when data are transmitted over noisy channel. Since standard data compression techniques are not designed for error correction, compressing data and transmitting over noisy channels may cause corruption of the whole compressed sequence. However, instead of employing standard compression techniques, like Huffman coding, data can be compressed using error correcting codes that are suitable for both data compression and error correction purposes. Recently Turbo codes repeat-accumulate codes and LDPC have been used as lossless source codes and have achieved compression very close to source entropy. When a near lossless compression is desired, i.e. a small level of distortion is acceptable, the source coder generates fixed-length codewords and the encoding complexity is low. This chapter introduces the main theorems that explain how the channel coding can be used as source coding and the intersection between data compression and error correcting codes.

2.1.1 Slepian-Wolf Coding (SWC)

Let $\{(X_i, Y_i)\}_{i=1}$ be a sequence of independent and identically distributed (i.i.d.) drawings of a pair of correlated discrete random variables X and Y For lossless compression with arbitrary small error ($X = \hat{X}$ and $Y = \hat{Y}$) after decompression. It is known from Shannon's source coding theory that a rate given by the joint entropy $H(X,Y)$ of X and Y is sufficient if they are jointly encoded (see Figure 2-1- (a)). For example, we can first compress Y into $H(Y)$ bits per sample, and based on the complete knowledge of Y at both the encoder and the decoder, we can then compress X into $H(X|Y)$ bits per sample. But if X and Y must be separately encoded for some user to reconstruct both of them the sufficient rate for lossless reconstruction is not known.

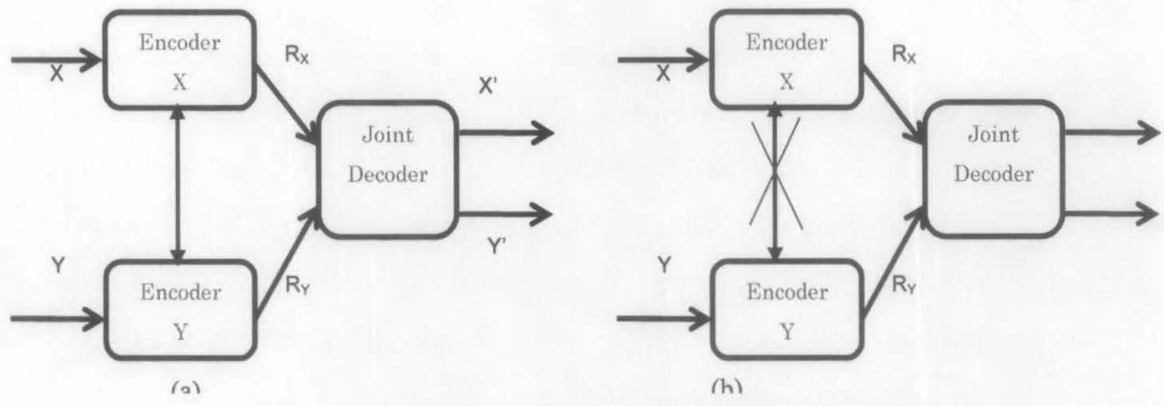

Figure 2-1 a) Joint encoding, b) separate encoding

One can separately encode them with rate $R = H(X) + H(Y)$, which is greater than $H(X,Y)$ if the two sources X and Y are correlated. In 1973 Slepian and Wolf [4] showed that $R = H(X,Y)$ is sufficient for lossless decompression, even for separate encoding of correlated sources (see Figure 2-1.(b)). By $R_X \geq H(X|Y)$,$R_Y \geq H(Y|X)$, and $R_X + R_Y = H(X|Y)$, which is shown on Figure 2-2.

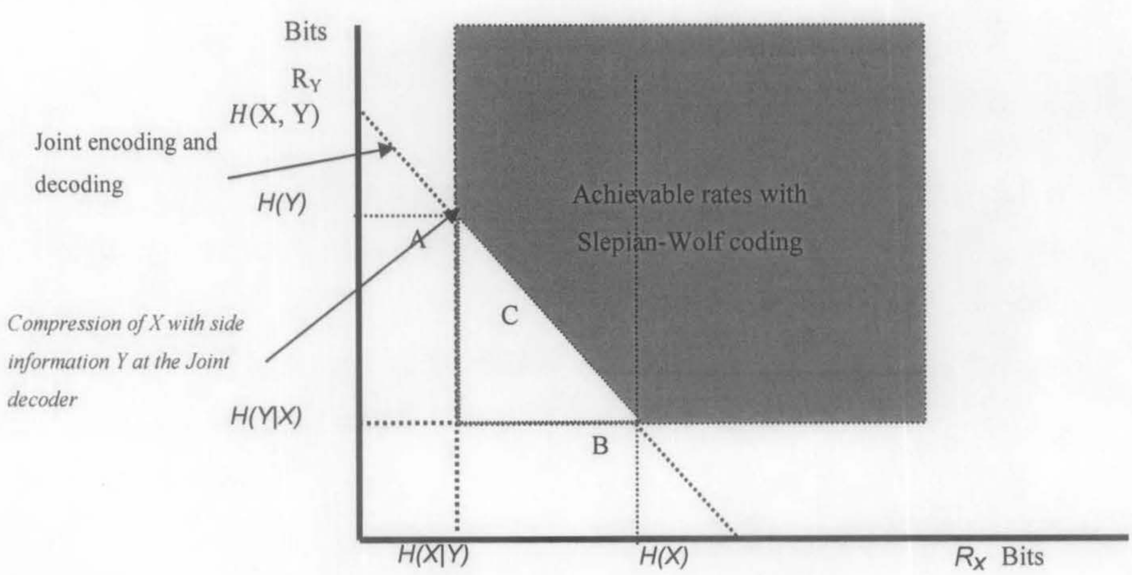

Figure 2-2 Achievable Slepian-Wolf range

2.1.2 Proof of Achievability of the Rate Region Specified By the Slepian Wolf Theorem:

A theorem of random bins is introduced to prove the achievability of the rates in the Slepian Wolf theorem [12, 18]. A large random index for each source sequence is chosen. If the number of these typical sequences is small enough, different source sequences will have different indices, and the source sequence is recovered from the index. Moreover this binning scheme does not require clear characterization of the typical set at the encoder; it is only needed at the decoder. This property is exploited and used for the distributed source case [18].

The basic idea of the proof is to partition the space of X^n into 2^{nR_1} bins and the space of Y^n into 2^{nR_2} bins, where $nR_1, nR_2 < n$.

Random code generation: Independently assign every $x \epsilon X^n$ to one of 2^{nR_1} bins according to distribution on $\{1,2,3 \ldots 2^{nR_1}\}$. Similarly, randomly assign $y \in Y^n$ 2^{nR_2} .

$$f_1 \colon X^n \to \{1,2,3, \ldots \ldots 2^{nR_1}\} \tag{2-1}$$

$$f_2 \colon Y^n \to \{1,2,3, \ldots \ldots 2^{nR_2}\} \tag{2-2}$$

Reveal the assignments to both encoders and the decoder.

Encoding: Source 1 sends the index of the bin which X belongs and source 2 sends the index of the bin to which Y belongs.

Decoding: Given the index of pair x_0, y_0 declare, $(\hat{x}, \hat{y}) = (x, y)$ if there is one and only one pair of sequence (x, y) such that $f_1(x) = x_0$ and $f_2(y) = y_0$ and $(x, y) \epsilon (X, Y)$ otherwise declare an error.

2.1.3 Wyner Scheme

Rate points A and B (see Figure 2-2), are referred to as an asymmetric code $(H(Y), H\ (X|Y))$, the rate point C $(H(Y|X), H(X|Y))$ is referred to as symmetric code. In asymmetric coding, the goal is to code X at a rate that approaches $H(X|Y)$ based on the conditional statistics of (or the correlation) model between X and Y but not the specific Y at the encoder, in other word reaching the corner point A (see Figure 2-2). Wyner realized the close connection of DSC to channel coding and proposed as a practical approach for Slepian-Wolf coding the use of linear channel codes [19]. As explained in the proof of achievability section the random binning is performed by partitioning the space of all possible source outcomes into disjoint bins (sets) that are the cosets of some "good" linear channel codes for the specific correlation model. Consider the case of binary symmetric sources and the Hamming distance measure; with a linear $(n; k)$ binary block code, there are 2^{n-k} distinct syndromes, each indexing a bin (set) of 2^k binary words of length n. For each bin is a coset code of the linear binary block code and Hamming distance properties of the original linear code are preserved in each bin. In compressing, a sequence of n input bits is mapped into its corresponding $(n - k)$ syndrome bits, achieving a compression ratio of $n : (n - k)$ [19]. This approach, known as "Wyner's scheme" [20] for some time, was only recently used in [19] for practical Slepian-Wolf code designs based on conventional channel codes like block and trellis codes.

Wyner's syndrome concept can be broadened to all binary linear codes if the correlation between X and Y can be modeled by a binary channel. Any state-of-the-art near-capacity channel codes such as Turbo [21] and LDPC codes [22] can be employed to approach the Slepian-Wolf limit.

If achieving the corner rate point A can be done, then the other corner point B of the Slepian-Wolf rate region can be approached by swapping the roles of X and Y and all points between these two corner points can be realized by time sharing – for example, using the two codes designed for the corner points 50% of the time each will result in the mid-point C. However, time sharing is not practical because it requires exact synchronization among encoders[23, 24]. Another method for achieving the midpoints is the source splitting approach [25], which has the potential to reach all points on the SW range by splitting two sources into three sub-sources of lower entropy. In[26] the system consists of two different Turbo codes which form a big Turbo code with four component codes. In a symmetric scenario(the rate of both encoders are the same), suggested half of the systematic bits from one encoder and half of the other are sent along with to the punctured parity bits from both encoders. The method for constructing a single code based on the syndrome technique proposed in [23], is to achieve arbitrary rate allocation among the two encoders. It constructs independent sub-codes of the main code and assigns them to different encoders. Each encoder sends only partial information about the source; by combining the received bit-streams a joint decoder will perfectly reconstruct the sources.

2.1.4 Asymmetric Codes Example

As an example, suppose we have two correlated 8-bit grayscale images X and Y whose same location pixel values x and y are related by x $\{y - 3; y - 2; y - 1; y; y + 1; y + 2; y + 3; y + 4\}$. In other words, the correlation of x and y is characterized by $-3 \leq x - y \leq 4$, or x assumes only eight different values around y. Thus, joint coding of x would take three bits. But in DSC, we simply take modulo of pixel value x with respect to eight, which also reduces the required bits to three. Specifically, let $x = 121$ and $y = 119$. Instead of transmitting both x and y at (8 b/p) without loss, we transmit $y = 119$ and $x' = x\,(mod\,8) = 1$ in distributed coding. Consequently, x' indexes the set that x belongs to, i.e.,$x = \{1; 8 + 1; 16 + 1; :::; 248 + 1\}$, and the joint decoder picks the element $x = 120 + 1$ closest to $y = 119$.

2.1.5 Wyner-Ziv Coding (WZC)

In communication applications, we are often dealing with continuous sources, then the problem of rate distortion with side information at the decoder arises. The problem of allocating the needed rate to encode X under the constraint that the average distortion between X and the coded version $\check{X}$ is$E\{d(X,\acute{X})\}$, assuming the side information Y is available at the decoder but not at the encoder (see fig (2.3)). First considered by Wyner and Ziv in [4], is one instance of DSC with Y available as side information at the decoder, corresponding to point A $(H(Y), H(X|Y)$ and B $(H(X), H(Y|X)$ (see Figure 2-2).

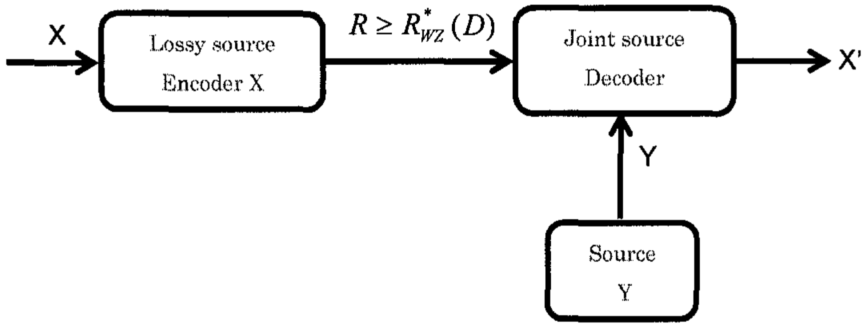

Figure 2-3: Wyner-Ziv coding or lossy source coding with side information

The Wyner and Ziv (WZ) work establishes the minimum rate $R^*_{WZ}(D)$ necessary to encode X guaranteeing its reconstruction with an average distortion below D, assuming that the decoder has the side information Y available. The results obtained by Wyner and Ziv indicate that when the statistical dependency between X and Y is exploited only at the decoder, the transmission rate increases comparing to the case where the correlation is exploited both at the encoder and the decoder, for the same average distortion, D. Mathematically, the Wyner and Ziv theorem can be described by:

$$R^*_{WZ}(D) \geq R_{X|Y} \rightarrow D \geq 0 \qquad (2\text{-}3)$$

Where $R^*_{WZ}(D)$ represent the minimum encoding rate for X and $R_{X|Y}(D)$ represents the minimum rate necessary to encode X when Y is available at the encoder and the decoder. This means if that the correlation between X and the side information is only exploited at the decoder, it is possible (theoretically) to reconstruct the sequence X with an arbitrarily small error probability.

Wyner and Ziv showed, that there is no rate increase, for all $D > 0$, when X and Y are jointly Gaussian sequences and a MSE distortion measure, this means that there is theoretically no rate loss in the transmission rate when the exploitation of the statistical dependency between X and Y is performed both at the encoder and the decoder. This case is of special interest in practice (example, sensor networks) because many image and video sources can be modeled as jointly Gaussian (after subtracting mean) [19]. A major work on practical Wyner-Ziv code design called DISCUS[12] recently extended the no rate loss condition for Wyner-Ziv coding to

$X = Y + Z$, where Z is independently Gaussian but X and Y could follow more general distributions. Generally, if the correlation between the sources output X and the side information Y can be modeled with a "virtual" correlation channel, then a good channel code over this channel can provide a good Slepian-Wolf code through the syndromes and the associated coset codes. Thus as Figure 2-4 depicts the apparently source coding problem of Slepian-Wolf coding is actually a channel coding one and near-capacity channel codes such as Turbo and LDPC codes can be used to approach the Slepian-Wolf limits.

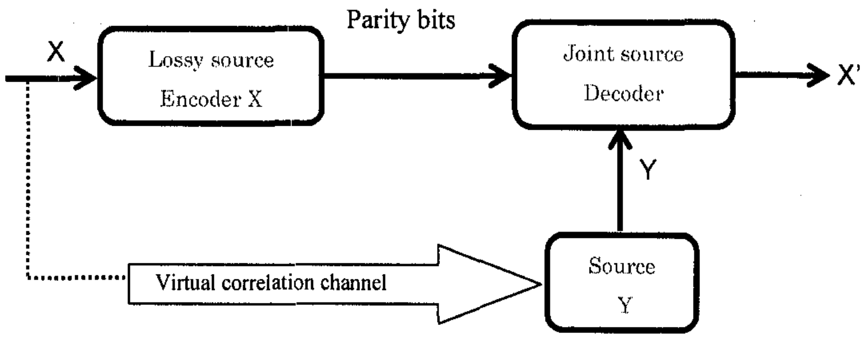

Figure 2-4: Relationship between channel coding and Slepian-Wolf coding.

Because distortion is introduced to the source in WZC, source coding (or quantization) is needed to quantize X. Typically there is still correlation remaining in the quantized version of X and the side information Y, and SWC should be employed to exploit this correlation to reduce the rate. Since SWC is based on channel coding, WZC is a source-channel coding problem [19]. There are quantization losses due to source coding and binning loss due to channel coding. In conclusion, the Slepian-Wolf and the Wyner-Ziv theorems suggest that it is possible to compress two statistically dependent signals in a distributed way (separate encoding, joint decoding) using a rate similar to that used in a system where the signals are jointly encoded and decoded like in traditional video coding schemes.

2.1.6 Successive Wyner-Ziv Coding Theory

The problem of successive refinement of information was originally formulated in [27]. An example of successive refinement might be image compression in which one briefly describes a coarse image and then follows with successive refinements of the description that further refine the image. The goal is to achieve the rate distortion bound at each stage. A source X is to be encoded and transmitted through a rate-limited channel. With rate R_1, the decoder produces X_1, which is an approximation of X as distortion level D_1. At a later stage, the encoder sends a secondary string at rate ΔR to the decoder. With both bitstreams at hand, the decoder will produce X_2, a more accurate reconstruction of X, at distortion level D_2. If successive coding in two or more stages can be made RD optimal simultaneously at all stages, the source is called *successively refinable*. For the two-stage case, the two rates should simultaneously lie on the R-D curve, i.e.,

$$R_1 = R^*_{X|Y}(D_1) \quad \text{and} \quad R_1 + \Delta R = R^*_{X|Y}(D_2) \tag{2-4}$$

where $R_X(D)$ is the RD function of the source X at distortion level D. It has been shown in [28] that a necessary and sufficient condition for a source to be successively refinable is that the conditional distributions $f(\acute{X}_1|X)$ and $f(\acute{X}_2|X)$ are Markov compatible in the sense that they can be represented as a Markov chain $X \rightarrow \acute{X}_2 \rightarrow \acute{X}_1$. A successive refinement code for the Wyner-Ziv problem consists of multi-stage encoders and decoders where each decoder uses all the information generated from the decoders of its earlier stages [28]. Figure 2-5 depicts a special case of two-stage successive coding for the Wyner-Ziv problem with the side information at each stage being the same.

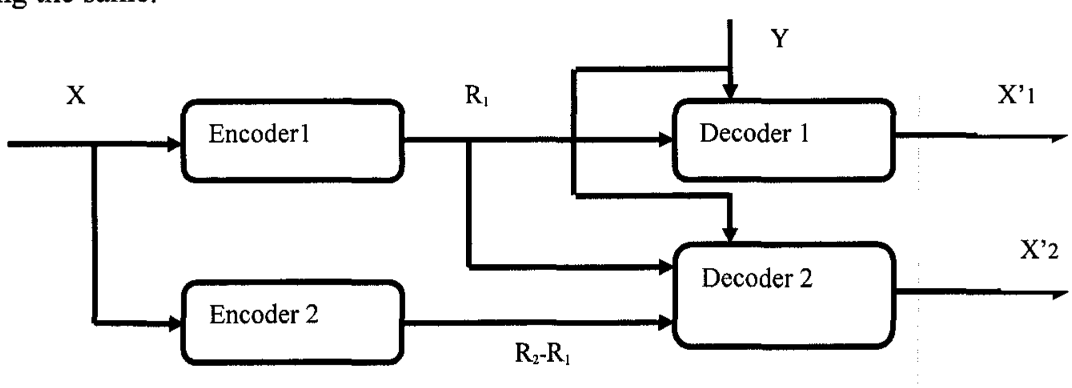

Figure 2-5 Two-stages successive refinement with identical side information at decoders

Let Y be the side information available to the decoder at both the coarse and the refinement stages and the corresponding coding rates (distortions) are $R_1(D_1)$ and $R_2(D_2)$, respectively. Let $R^*_{X|Y}(D_1)$ be the Wyner-Ziv RD function [5]. According to (2.4), a source X is said to be successively refinable from D_1 to D_2 $(D_1 > D_2)$ with side information Y if (equation 2-4) is satisfied.

The notion of successive coding can be naturally extended to any finite number of stages [28].Consider the case when the side information fed into the K decoders at each level is the same, the source X is multi-stage successively refinable with side information Y if

$$R_1 = R^*_{X|Y}(D_1) \quad \text{and} \quad R_i + \Delta R_i = R^*_{X|Y}(D_{i+1}) \quad \text{for } i = 1,2,\dots.k-1 \qquad (2\text{-}5)$$

Necessary and sufficient conditions for successive refinability are given in [28] and the jointly Gaussian source (with MSE measure) shown to be multi-stage successively refinable in the Wyner-Ziv setting

2.1.7 Channel Coding

Several commonly used channel codes are used in Slepian-Wolf codes. These codes show superior error correction performance in communication systems against noisy channel conditions. A Slepian-Wolf code has a better compression performance if its core, the channel code, has better error correction ability. Therefore, some state-of-the-art channel coding techniques are reviewed in this section.

Turbo codes is high performance forward error correcting codes and other generally concatenated codes are a class of near-capacity channel codes that use concatenation of two encoders cross random interleaving in the encoder and iterative maximum-a-posteriori (MAP) decoding, also called Bahl-Cocke-Jelinek-Raviv (BCJR) [29]decoding, for the two component codes in the decoder. A carefully designed concatenation of two or more codes performs better than each of the component codes alone [30]. The role of the pseudorandom interleaver is to make the code appear random while maintaining enough code structure to permit decoding [21]. The BCJR decoding algorithm allows exchange of information between decoders and thus an iterative interaction between the decoders (hence Turbo), leading to near capacity performance.

LDPC codes [22] are block codes, described by their parity-check matrix H and the associated bipartite graph. The parity-check matrix H has a small number of one's hence low densities. The degree of distribution polynomials $\lambda(x)$ and $\rho\ (x)$ describe how these one are spread and define the percentage of columns and rows of H respectively, with different Hamming weights (number of ones). When both $\lambda(x)$ and $\rho\ (x)$ have only a single term, the LDPC code is regular, otherwise it is irregular. In general, the irregular LDPC code is expected to be more powerful than a regular one of the same code word length and code rate. The code rate is exactly determined from $\lambda(x)$ and $\rho\ (x)$.

The bipartite graph is used in the decoding procedure, to allow the application of the message-passing algorithm [22]. (Iterative decoding algorithms including the sum-product algorithm, the forward/backward algorithm in speech recognition and the belief propagation algorithm in artificial intelligence are collectively called message-passing algorithms.) LDPC codes are the most powerful channel codes nowadays [22].

Both Turbo and LDPC codes perform near-approaching performance with more complexity than most conventional channels; and algorithms have been proposed for the design of very good Turbo and LDPC codes. The design procedure for LDPC codes is less complex and therefore allows faster, easier and precise design. Which has made LDPC codes the most powerful channel codes over both conventional and unconventional channels. But Turbo and more generally concatenated codes are still employed in many applications when short block lengths are needed or when one of the component codes is required to satisfy additional properties (e.g., for joint source-channel coding) [21, 22].

2.2 Basic Wyner-Ziv Coding Architecture

The WZ coding involves three elements; the signal processing (source coding), communications (channel coding), and estimation theory. Video coding scheme based on the WZ coding system is known as Wyner-Ziv video coding. Such system architecture encloses three main modules; each one to perform a single function of the WZ coding elements. In traditional video coding schemes, the X and Y sources can be, for instance, the odd and even frames of a video sequence respectively. The following is a simplified walkthrough of the basic Wyner-Ziv coding architecture as shown in Figure 2-6.

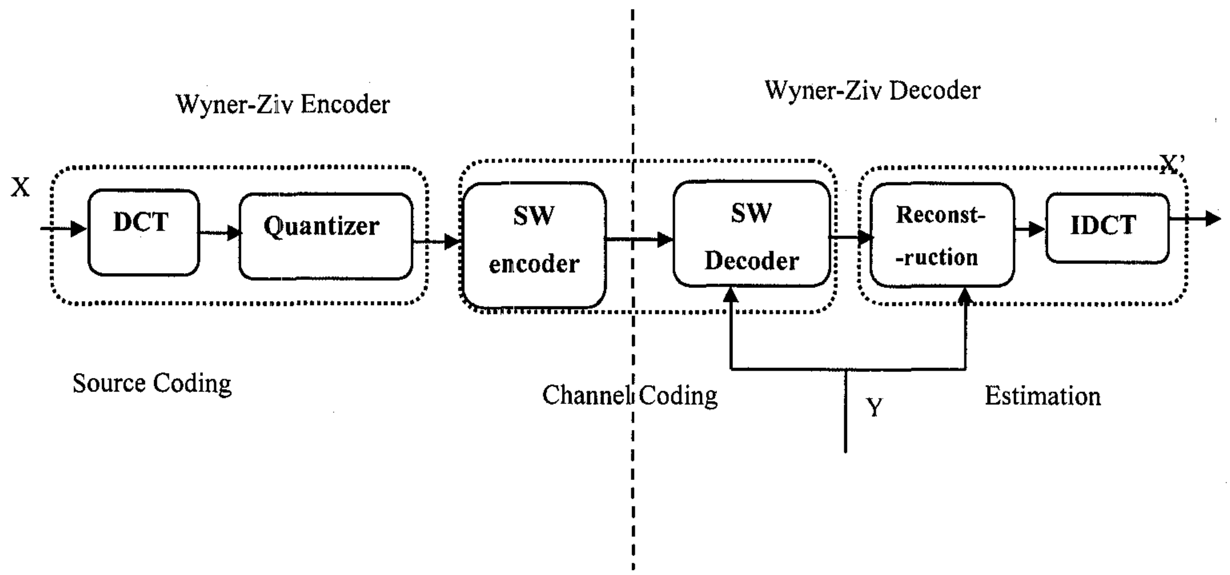

Figure 2-6: Block diagram of the basic Wyner-Ziv codec

2.2.1 Transforming in the Basic Wyner-Ziv Coding Architecture

In the case of implementing the DVC coding in the transform domain the first stage to encode the main information is the transform coding. Typically, transform coding is applied over $n \times n$ sample blocks of a frame to de-correlate those frame samples. The result of that de-correlation process is the block energy concentration in a few large valued transform coefficients; these coefficients are called low frequency transform coefficients because they typically represent lower frequencies (closer to DC). In the conventional video coding to reduce the number of bits required to encode X by transmitting only the low-frequency coefficients to the decoder (without this shortcut being perceptible in the reproduced frame quality due to the HVS limitations), since those coefficients contain most of the block information and have the largest values. In[31] it was determined for the coding scenario that the discrete cosine transform

(DCT) is an optimal choice for the transform module, in terms of rate-distortion performance. The results obtained confirm the main idea that exploiting the spatial correlation within a frame at the encoder leads to a rate-distortion improvement. This improvement is measured over a pixel domain scheme where just quantization and Slepian-Wolf coding is used to encode the sequence as shown in Figure 2-6.

2.2.2 Quantizer (Source Coding)

The source X is the pixel values of the even frames. The samples of the sequence X are first quantized using a uniform quantizer. The quantization compresses a range of values in the source into a single value and the space of a signal is divided into intervals called bins. When the number of bins is less than the total number of the values a signal can take, the total bitrate is reduced. With the purpose of reducing the bitrate, the quantization stage is also applied to DVC. The conditions of a quantizer that suits the DVC were studied in [31] where it studied the optimality conditions which such quantizers must satisfy, together with an extension of the Lloyd algorithm for a locally optimal design. Also the work in [31] provided a theoretical characterization of optimal quantizers at high rates and their rate-distortion performance, and applied it to develop a theoretical analysis of orthonormal transforms for distributed coding. Bits are extracted from the quantized symbols (q) and grouped together to form bitplane starting from the most significant bit (MSB) toward the least significant bit (LSB).

2.2.3 Slepian-Wolf (SW) Codec (Channel Coding)

The goal of the Slepian-Wolf codec is to reduce the transmission rate of the main information so it can approach Slepian-Wolf limits. This process is performed by channel coding techniques. The Slepian-Wolf encoder module plays the role of generating redundant information (parity bits) as in channel codes. The Slepian-Wolf encoder is fed by the bitplanes extracted from the quantized symbols. The Slepian-Wolf encoder outputs the systematic bitplane and the parity bits. The systematic output is discarded and only the parity bits are transmitted. The idea is that, Y being the noisy version of X, the parity bits can correct this noisy estimate Y. In the basic Wyner-Ziv decoder, the quantized symbol stream is decoded through joint source-channel decoding with the aid of the sequence Y, hereafter also called the side information. Practical Slepian-Wolf coding is implemented by using practical channel coding techniques. In the next section three practical Slepian-Wolf coding methods will be discussed.

1) Syndrome based Slepian-Wolf coding

The distributed source coding scheme called "Distributed source Coding Using Syndrome" (DISCUS) was proposed by Pradhan and Ramchandran in 1999. In DISCUS, the codeword space is divided into several non-overlapped cosets. The coset contains a group of codewords, the minimum Hamming distance between these codewords kept as maximum as possible. The encoding of source information X is done by sending only the index of the cosets that contains X. The decoder receives the coset index and decodes it into one of the codewords within the cosets. The codewords are chosen as the closest in Hamming distance to the side information. The encoding objective is to design a codebook for the source information X to optimally partition the source codeword space into a channel code cosets.

2) Turbo codes based Slepian-Wolf coding

Several practical implementations have been independently proposed for the Slepian-Wolf coding all based on Turbo codes [11, 14, 32, 33]. In [32] two identical concatenated finite-states machine (FSM) and separated by an interleaver are used as a Slepian-Wolf coding core. Both FSM encoder output n for every k binary input symbols, $k>n$, and the total rate of the parallel concatenated encoder is k/2n. To achieve the points on the straight line between the two corners in the Slepian-Wolf achievable range, punctured Turbo codes are proposed for the SW coding [32]. The proposed framework encloses two Turbo encoders connected by an interleaver which is considered as a super Turbo encoder, with two different but correlated input sequences. Another popular scheme based on rate-compatible punctured Turbo codes (RCPT) is proposed by Aaron[11]. In this proposed system a systematic Turbo encoder is used; however the systematic bits are discarded. Only a subset of parity bits is sent to the decoder which is determined and requested via using feedback channel by the decoder. The decoder uses the received parity bits and the side information to recover the source information.

3) Low-density parity check codes based Slepian-wolf coding

Low density parity check codes (LDPC) have the ability to achieve near-capacity performance when iterative message-passing decoding and sufficiently long block size is used [19] several Slepian-Wolf coding methods have been proposed based on the LDPC [23, 24, 34]. Work in [23], presented a distributed linear block code. This single block code allows attaining any point on the Slepian-Wolf achievable rate region. Its achievable region is smaller than theoretical the Slepian-Wolf achievable region, because LDPC codes cannot reach Shannon limit. In [19], systematic irregular repeat-accumulate (IRA) is used as a Slepian-Wolf coding [35] method. The IRA encoder generates the syndrome bits and sends them to the decoder. The decoder uses a message-passing algorithm to decode from the received syndrome bits and the side information. Performance comparison made in [35] showed that Slepian-Wolf coding based on IRA LDPC codes achieves a rate closer to the Slepian-Wolf bound than Slepian-Wolf coding based on Turbo codes. To overcome the disadvantage of specifying the code rate before processing by LDPC codes, a rate adaptive Slepian-Wolf coding algorithm was proposed in [5, 11, 16, 34-41]. The algorithm uses accumulate LDPC codes, the LDPCA encoder consists of an LDPC syndrome-former concatenated with an accumulator. The LDPCA decoder performs rate-adaptation by modifying its graph each time it receives an additional increment of the accumulated syndrome.

2.2.4 Reconstruction (Estimation)

In this final stage of reconstruction (estimation), an estimate of the main information X is obtained using the decoded $(\acute{q})$ quantized symbols and the side information Y. In the DVC framework, the optimal reconstruction $\acute{X}$ is the value that minimizes the distortion $d(X, \acute{X})$. In [3], the reconstruction is achieved by maximizing the conditional expectation ((2-6) by choosing the appropriate $\acute{X}$ so as to reconstruct the source information X (pixel value or DCT coefficient)

$$E(X|q,y) \quad (2\text{-}6)$$

assuming a Laplacian distribution to model the residual between corresponding elements of X and Y (side information).In other works [6, 9], a linear estimate algorithm is used to reconstruct the DCT coefficients as in (4)

$$\acute{x} = a_0.\acute{q} + a_1.y \quad (2\text{-}7)$$

where a_0 and a_1 are obtained offline from certain training procedure.

In the literature of Wyner-Ziv video coding some modules have been added to the basic Wyner-Ziv Codec architecture depicted in Figure 2-6. These added modules aiming to enhance the overall performance of the systems, in the remaining subsections these modules and their function will be explained.

2.2.5 Side Information Generation Process

The distributed video coding solution is developed based on the coding with side information principle. The performance of the distributed video coding system is fundamentally determined by the quality of the side information [11, 16, 36-41]. Two factors characterize the efficiency of the video coding system the rate and the distortion. The influence of the side information quality in terms of rate performance is noticeable, because if the predicted frame is very similar to the original frame few coding errors need to be corrected. Therefore less parity bits have to be sent to the decoder, and better overall compression can be achieved. But if the side information generation fails a lot of errors have to be corrected by the parity check bits which results sending more bits from the encoder and degrading the compression performance of the DVC system. The quality of the side information also affects the distortion on the reconstructed frame. Since the reconstruction process of the decoded frame incorporates the side information and the quantized symbol to reconstruct the original frame. Consequently if the generated side information is very similar to the original frame a good quality reconstructed frame can be obtained. But if the side information generation process generates a poor prediction of the original frame, the quality of the reconstructed frame is degraded.

In goal to achieve better coding efficiency in terms of rate-distortion, by making available at the decoder, a better estimate of the side information from temporally adjacent frames, an efficient frame prediction tool is studied in many studies. Two major approaches can be identified in the literature:-

1. The trajectory based motion compensation.
2. The hash based motion estimation.

The trajectory based motion estimation approach completely adheres to the low complexity encoder constraints, which is main difference between the distributed video coding and the conventional video coding. The encoder does nothing to help the decoder estimate side information; which requires the decoder part to fully exploit the temporal redundancy between adjacent frames and handles the motion estimation and compensation. The motion estimation is performed carefully in order to ensure that only true motion is captured and that a reliable predicted frame is produced. The traditional motion estimation and compensation techniques used at the encoder for hybrid video coding are not adequate to perform motion estimation since they attempt to choose the best prediction for the current frame in the rate-distortion sense [42]. For frame prediction, we need to find the best estimate frame for the original frame, and therefore a good criterion is to estimate true motion, and based on that to perform motion compensation between temporally adjacent frames. In this approach since the original frame is not available at the decoder, it is common to make assumptions about the type of motion in the sequence. Motion field research has provided many useful insights into the reconstruction of true motion information without the original video sequence. For example, linear motion model can be used, in which it is assumed that the motion in the current frame is a continuous extension of the previous frames' motion. While this is true for some video sequences, the motion of natural video sequences is not well defined and a simple model can be inadequate [43]. Several implementation methods were proposed for the trajectory based motion estimation approach. They can be divided into two classes, the interpolation based method and the extrapolation based methods. The first method uses the previous and the next decoded frame as references to interpolate the side information. The reference frames in a typical DVC scenario are key frames (GOP=2), but they could also be previously reconstructed Wyner-Ziv frames (GOP > 2). The simplest interpolation that can be used is to make the side information equal to the previous temporally adjacent frame which means to assume that there is no temporal variation [38], or to perform bilinear interpolation between the key frames (GOP=2). However, use of these techniques to generate the side information in medium or high motion video sequence (*foreman, sequences*) will produce rough estimate since similarity between two temporally adjacent frames is low. A rough estimate will degrades the overall performance of the system in terms of rate-distortion as we mentioned earlier. Experiments also show that also these simple techniques will introduce some "jerkiness" and "ghosting" artifacts in the reconstructed frames [42]. Powerful interpolation techniques are sought based on the motion estimation/compensation techniques; the motion estimation techniques have already been used in the conventional video coding and known to be very effective tools in the motion estimation-compensation problem. Typically these principles attempt to find a coherent motion to express the true motion in the sequence[40]. Some proposed techniques to implement motion estimation-compensation principle include 3D recursive blocks matching in an attempt to find the true motion [44-48]. The Simplest estimator to generate side information using the motion estimation/compensation principle is depicted in figure 2-7. First, the previously decoded key frames X_{t-1} at time $(t - 1)$ is used as a reference and next reconstructed key frame X_{t+1} at time $(t + 1)$ is used as source to calculate the forward motion vector MV_F. Then it uses the next reconstructed key frame X_{t+1} as reference and the previously decoded key frames X_{t-1} as source to calculate the backward motion vector MV_B. After that, it applies $\frac{MV_F}{2}$ on the previously reconstructed frame X_{t-1} to generate a frame P_F and $\frac{MV_B}{2}$ on X_{t+1} to generate a frame P_B. In fact, it models the current frame X_t as half way between frames X_{t-1} and X_{t+1} [13]. However, as the current frame is not available to decide which motion vector is best, $\frac{MV_F}{2}$ and $\frac{MV_B}{2}$, are both averaged.

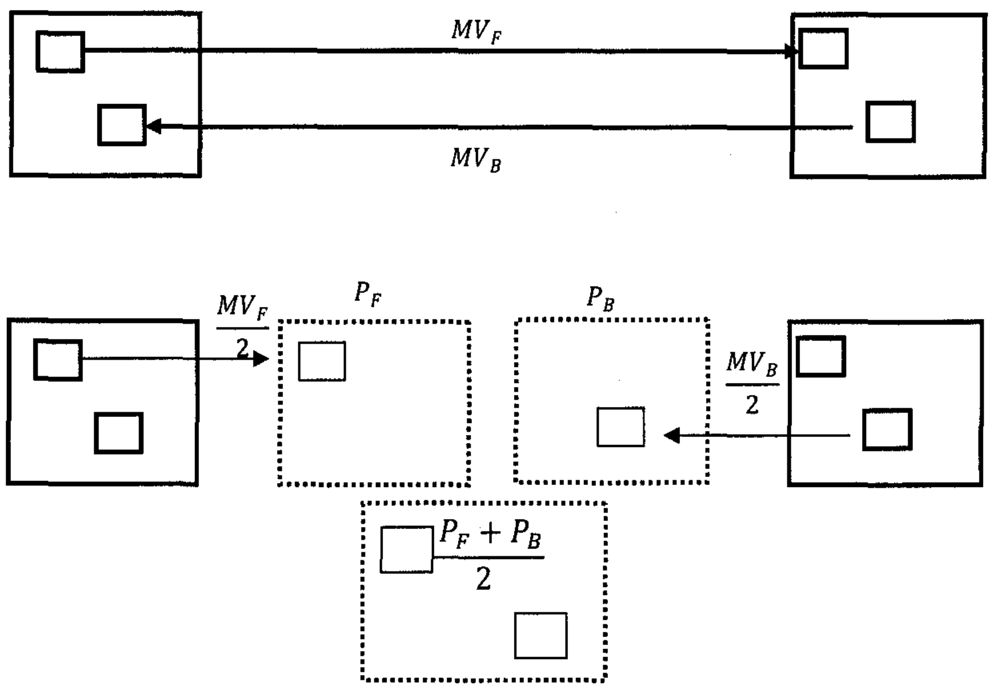

Figure 2-7: Side information generation

After the motion estimation and compensation generally there will be three cases for a given pixel

- It is uniquely defined by a single motion vector.
- It is defined by more than one motion vector (an overlapping occurred).
- It is not defined by any motion vector (it is left blank).

The second and third cases require smoothness constraints for the estimated motion and overlapped block motion compensation as well as a solution for the left blank areas [40]. A spatial smoothing algorithm motivated by the observation that the motion vectors have sometimes low spatial coherence; proposed in [42] targets the reduction of the number of false motion vectors. The proposed smoothing algorithm uses a weighted median filter to maintain the motion field spatial coherence by looking at each block, for candidate motion vectors at neighboring blocks, see Figure 2-8. This proposed spatial smoothing algorithm is both effective at the image boundary, where abrupt changes of the direction of the motion vectors occur, as well as in homogenous regions where outlier are effectively removed. The weighted median vector filter is defined as (2-6)

$$\sum_{j=1}^{N} w_j \left\| x_{wvmf} - x_j \right\| \leq \sum_{j=1}^{N} w_j \left\| x_i - x_j \right\| \tag{2-6}$$

Where $(x_1, x_2, \dots\dots x_N)$ are the motions vectors of the current block in the previously interpolated frame and the neighboring blocks; $w_1, w_2, \dots\dots w_N$ correspond to the set of adaptive-varying weights and w_{wvmf} represent the motion vector output of the weighted median filter.

N1	N2	N3
N4	B	N5
N6	N7	N8

Figure 2-8: Neighboring blocks of block current for weighted median vector filter

In [42] first the reference frames are low pass filtered to improve the reliability of the motion vectors; to help estimate motion vectors closer to the true motion field, the forward and backward motion vectors estimation is performed on filtered frames. Then, a spatial smoothing process is applied as depicted in Figure 2-9.

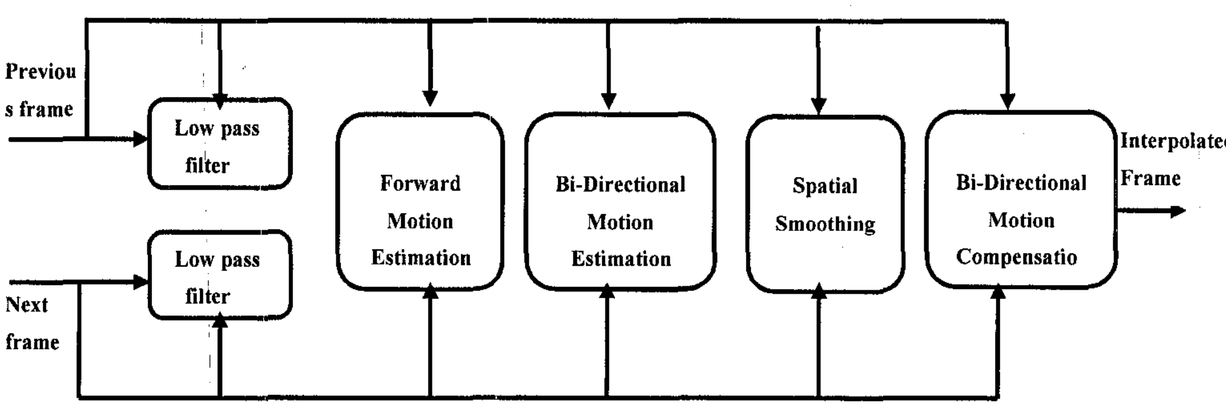

Figure 2-9: Proposed frame interpolation framework [5]

A simple solution to the case where a pixel is not defined by any motion vector (left blank), was proposed in [49] where after the averaging of P_F and P_B there is enough information available about the current frame to perform motion estimation using the averaged frame and the previous frame X_{t+1}. The averaged frame is divided into blocks 16x16 pixels, and then if there is a blank area in a macroblock, motion estimation will be performed in the macroblock. The black area will not be considered when calculating the sum of absolute difference (SAD), once a new reference block is found in frame X_{t+1}, that black best represents the current information we have about that area. So the pixel in the reference frame X_{t+1} will fill the blank area in the current macroblock. To improve the method, bi-directional motion estimation is performed to fill the blank area and a reference block will be searched for in both the previous frame X_{t-1} and the next frame X_{t+1}.

Although the above frame interpolation techniques perform well, they have several limitations. Specifically the displacement vectors are estimated using a block with no semantic meaning, which prevent the coherent capture of the moving objects in the scene [49]. Moreover, the fixed-size blocks are usually inefficient for large static background and areas with sharp moving edges. Additionally, the performance of the block matching algorithm BMA with fixed-size blocks relies on the assumption of uniform motion within blocks [11, 38].

A frame interpolation technique motivated by the inadequacy of the fixed-size blocks motion estimation method is proposed in [49]. The proposed technique uses the adaptive block size and adjusts dynamically the block size based on motion analysis. Ideally the adaptive block size should adapt to motion characteristics, the proposed method initially perform motion estimation using the fixed-size block, then it merges areas with similar motion in variable sized blocks , additionally a block splitting is used to define smaller blocks for motion estimation. The splitting is based on checking the direction of motion vectors in a 3x3 window of 8x8 blocks; if block merging has not been applied in the sliding window and the directions of neighboring motion vectors are classified to more than four directions the central block is split into four 4x4 blocks. Finally when motion vectors of variable sized block have been estimated, bi-directional motion estimation is applied to estimate the interpolated frame.

The idea of using the spatial correlation to enhance the motion estimation-compensation based for interpolation has also been considered in the pixel domain DVC system as proposed in [38]. The frame to be encoded X_t is divided into two subsets (X_t^A, X_t^B), shown in Figure 2-10, based on coordinates of the pixels. Then, the encoder processes the two subsets independently. The side information is obtained by motion interpolation of interpolated frame X_{t-1} and X_{t+1} as described in [42]. The same dividing process is performed on the interpolated frame Y. Divided into two subsets (Y_t^A, Y_t^B) the first interpolated subset is corrected by the decoder. Then, the next is step where the spatial correlation is performed, in which spatial side information is obtained by interpolating pixel values in subset X_t^A in the pixel locations in subset Y_t^A.

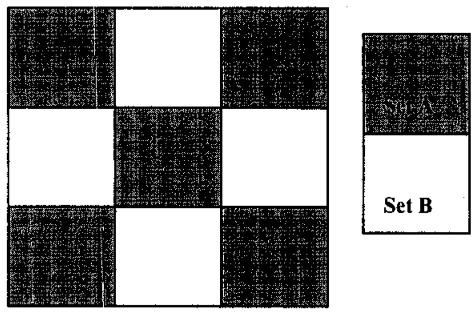

Figure 2-10: Pixel neighborhood used for interpolation

The second method using the trajectory based motion compensation approach is the extrapolation methods i.e. using the previously reconstructed frames to predict the forthcoming frame at the decoder. This method is motivated by the observations that the advantages of interpolation is future information about the uncovering areas, but it only holds if the temporal distance is small enough ($GOP = 2$). But keeping the temporal distance small has drawback of either a high bitrate or low quality for the intracoded frame in case of low bitrate. In contrast motion estimation and compensation in extrapolation case can be done over a temporal distance of only one, regardless of the GOP-size. The main advantages of extrapolation are the sequential decoding for low latency [38, 50, 51], as well as possible bitrate saving due to a less intracoded frame. Because, in the case of a successful decoding it is possible to use the previously decoded WZ-frame.

The simplest extrapolation technique is to use a previously reconstructed frame. However since the motion is generally present in video sequences, an extrapolation based on the motion observed in previously reconstructed frames fits better the purpose of producing an extrapolated frame similar to the WZ-frame being decoded, by simply the projecting the observed motion in the previously reconstructed frames to the WZ-frame time slot. In [51], using the same interpolation enhancement tools (motion estimation and spatial smoothing) but different reference frames is proposed. First, motion estimation is performed by using previously reconstructed frames; the block overlapping is used to reduce the block artifacts in the extrapolated frame. Then, it performs some motion field smoothing, in which, for each block a new motion vector is calculated by averaging all neighboring motion vectors. It projects pixels of the last reconstructed frames the motion vector obtained earlier; assuming that the motion is linear there warping of frame X_{t-2} into X_{t-1} will linearly continue from frame X_{t-1} to frame X_t. The proposed method also performs averaging in case of the overlapping (Occluded areas) and spatial interpolating in case of left blank areas (De-occluded areas). In [38] another extrapolation technique is proposed, using the same processes as [51]. It only differs in filling the blank areas by assuming that they always belong in the background where copying the neighboring pixels can be a perfect filling to these areas.

The distributed video coding main objective (low-complexity encoder) is completely achieved by using the trajectory based motion compensation approach. But due to its limitations; notably the difficulty of interpolating frames when high motion occurs, when the reference frames are temporally far from each other (long GOP size) and when coarse quantization is applied in the reference frames. These limitations have motivated a second approach the hash based motion estimation. The hash based approach simply requires the encoder to participate and send some information to the decoder to help in producing the side information generation. The problem of extracting some information to help in generating the side information is a tradeoff between extracting sufficient information and lowering the complexity of the encoder, the most sufficient information is the motion vector as in the case for the conventional video coding systems, but avoiding the extracting of these motion vectors is the main objective of the DVC. The first proposed DVC system (PRISM) [9] is a good example for this approach; where the encoder generate the 16-bit cyclic redundancy check (CRC) for the quantization indices of the first 16 low frequency coefficients of the inter-coded block. The decoder interpolates estimation for the frame being decoded, generates the 16-bit CRC for the quantization indices of the first 16 low frequency coefficients and compares it to the CRC transmitted by the encoder. If the CRCs match, the motion search terminates. If the CRCs do not match the decoder generates another candidate. The process of hash extracting in this system is much simpler than the motion vectors extraction which preserves the objective uncompromised. For the bit based DVC approach also the use of some help from the encoder is proposed in [1, 36, 52]. In the bit based DVC approach, the side information is more important than in the symbol based DVC approach because the decoder creates a final estimation and uses it in the decoding and reconstructing processes, so poor estimation would lead to poor performance. To accomplish the goal of sending help information to the decoder it is necessary to find the transform function (hash is block based) that maximizes the side information quality with the lowest possible rate. In [1], the essential properties for the transfer function were stated as *i*) energy compaction: the transform should generate a hash which can be compacted and represented with low amount of bits; ii) Robustness; iii) collision avoidance, each block has a unique hash; a final constraint is on the complexity of the transform function which should be as low as possible in order to maintain the low complexity encoder.

In [36, 52] the hash code for the frame block simply consists of a very coarsely sub-sampled and quantized version of the block. Since the hash is much smaller than the original data, the encoder is allowed to keep the hash codewords for the previous frame in a small hash store. For each block of the current Wyner-Ziv frame, the distance from the corresponding robust hash of the previous frame is calculated. With the current hash code, this distance is the squared difference between the sub-sampled version of the current block and the collocated quantized samples from the previous frame. If the distance is smaller than a threshold, a "no hash bits" codeword is sent. If the distance exceeds the threshold, the hash of the block is sent, along with the Wyner-Ziv bits. The proposed hash based technique in [1] is similar to [36, 52] in the sense that subset of the original DCT coefficients is sent to the decoder to guide the motion estimation process. It differs in adaptively selecting which DCT bands are necessary to be sent and which quality should be used to encode the selected DCT coefficients. The sent bands in [1, 36, 52] start with DC bands and since the block are very small, the DC coefficient is very important to distinguish each block in a robust way, especially for low bitrate hash where the hash rate must be kept small. In [1], the AC bands are selected based on the energy in the corresponding AC bands. In hash design, there are several matching criteria to measure the similarity between two blocks; the most straightforward one is to perform inverse quantization of the DCT hash coefficients to obtain a coarsely reconstructed frame to guide the motion estimation, in this case the SAD criteria between the reconstructed and candidate pixel blocks should be minimized. Another possibility is to calculate DCT hash for each candidate block and compare it with the DCT hash received, in this case ME is performed in a DCT domain where the SAD between the received hash and the calculated one for each candidate block is minimized.

2.2.6 Virtual Channel Modeling

Distributed source coding relies heavily on efficient error correcting codes; the performance of these codes depends greatly on the choice of the noise model that characterizes the dependency channel [53]. The no loss result of the Slepian-Wolf theorem comes under the assumption that the statistical dependence between the WZ data and the side information is perfectly known to both encoder and decoder, and that it follows a Gaussian distribution [54]. Exact knowledge of the statistical dependence between X and Y is required, 1) to characterize the channel in the SW decoder, 2) to perform an MMSE estimation in the inverse quantizer, and 3) to help in controlling the SW code rate[16]. It is common in DVC literature [1, 3, 11, 54-56] to use a Laplacian distribution to model the statistical correlation between the original frame and the side information (see Figure 2-11& Figure 2-11), this Laplacian distribution is used to convert the side information (pixel or DCT coefficient values) into soft-input information needed for channel decoding. In traditional video coding, the Laplacian distribution is typically used to model the distribution of the motion-compensated residual DCT coefficients. More accurate models can be found in such as generalized Gaussian distribution, however the Laplacian is chosen as a good tradeoff between model accuracy and complexity [57].

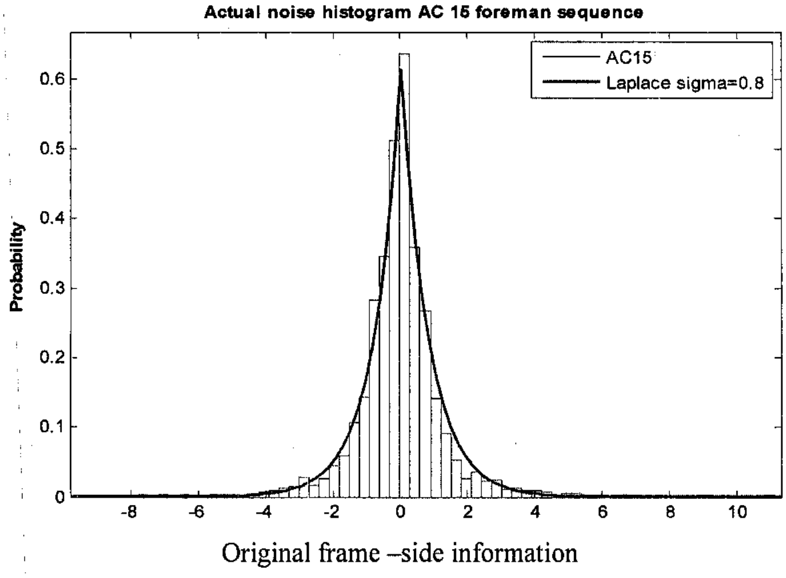

Figure 2-11: Actual noise histogram for AC band 15, foreman sequence [57]

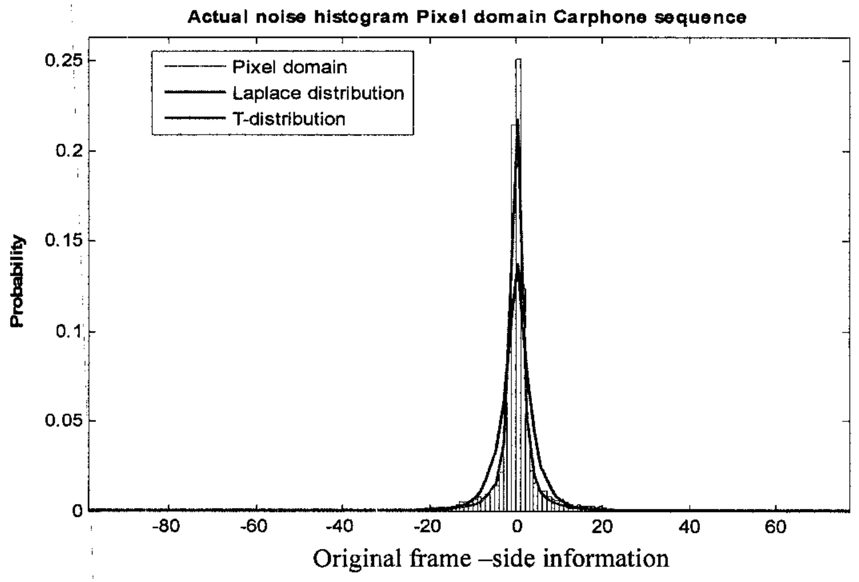

Figure 2-12: Actual noise histogram Pixel domain Carphone sequence [57]

In the first DVC implementations, the Laplacian parameters were off-line computed for each sequence. In [57], the correlation between the decoder side information and the original WZ frame is estimated at the encoder by recreating the side information for each WZ frame as the average of the two temporally closer key frames. Furthermore, the bit error probability of each bit plane is modeled assuming a Binary Symmetric Channel (BSC). In the SEASON framework [58], the deviation of the side information from the actual video frame is modeled as an additive stationary white noise signal. In regions where the motion estimation completely fails, caused by occultation, the side information is uncorrelated with the original data; therefore the noise process then models the effect of occlusions [16]. This process will have a large variance and follows, for instance, a uniform PDF model. In the non-occluded regions the noise process is assumed to be well-behaved, and follows for instance a Gaussian or Laplacian PDF model with a small variance [16]. In [53], with the use of Turbo codes as Slepian-Wolf coder, an inaccurately chosen noise model of the dependency channel (modeled by the conditional PDF $P(X|Y)$ has a big influence on the performance of Turbo decoding. The performance loss due to the inaccuracies in modeling the non-stationary dependency channel and the sensitivity of LDPC codes to the modeling of the noise in the dependency channel experimentally has been studied in [53]. Work in [59] has shown that for most of video resources, the actual AC coefficient distribution is closer to a Cauchy distribution than Laplacian distribution; also shown is that formulation of the relation between the quantizer and the output bitrate and distortion caused by the quantization parameter is more accurate by using Cauchy *pdf.* In [59] it was concluded that no one density function can be used to approximate each coefficient. It was also stated that Laplacian distributions provide the best fit for most of the coefficients, but if all coefficients must be summarized by one density function a Cauchy distribution is best. The work in [60] concluded that DCT coefficients follow neither Cauchy laws nor Laplace laws but mixtures of 1,2,3 Gaussian laws, their result were confirmed by comparison between coefficients and theoretical law probability densities. [61] also introduced that the general Gaussian distribution provides a better fit.

2.2.6.1 Estimation of Channel Model Parameters

The correlation noise can be interpreted as a virtual channel with an error pattern characterized by some statistical distribution (or model) since side information may be seen as a "corrupted" version of the original information. The virtual channel in the Pixel domain corresponds to the residual between the pixels of the original information and side information frames. While it corresponds to the residual between corresponding DCT bands of the WZ and SI frames in the Transform domain scenario. Since the SI quality varies along time and space, the choice of the unit at which the model parameters is computed influences the overall performance of the DVC coding system, using the video sequence as a unit can be considered as coarse estimate, a finer estimate can be obtained computing by using a smaller unit such as frame or block, or may be an even finer estimate such as using the pixel as a unit. The virtual channel model is given by:

$$p[WZ - SI] = \frac{\alpha}{2}\exp[-\alpha - |WZ - SI|] \qquad (2\text{-}9)$$

Where *p(.)* is the probability density function and α is the Laplacian distribution parameter defined by

$$\alpha = \sqrt{\frac{2}{\sigma^2}} \qquad (2\text{-}10)$$

α is varying temporally different α values for each frame, even for finer unit α values can be computed for each block and pixel.

There are two approaches to characterize the correlation model, offline and online; the first one is used in most of the literature [3, 9, 62]. The offline approach assumes spatially and temporally stationary module and several video sequences are used for training to obtain the model parameters [3]. This approach obtains the model parameters at the encoder, so the encoder performs the side information estimation. The offline approach, beside it increasing the complexity of the encoder it also assume the stationary model for all video sequences frames, which is not true for natural video sequence. The online approach performs the model characterization process at the decoder and it does not assume stationary model, in addition it can also be obtained for finer granularity (frame, block, pixel, and coefficient), which results in a better RD performance [57]. The model characterization requires the original frame to be available at the decoder which is not possible. Novel online estimation techniques are proposed in[57], where key frame are used instead of the original frames to obtain the model parameters.

2.3 Practical Wyner-Ziv Video Coding Approaches

Two groups have been working on the development of practical distributed video approaches. Kannan Ramchandran's group at the University of California is working on the block level and Bernd Girod's group at University of Stanford working on the frame level. In the following subsections, these two solutions will be briefly presented. This section also includes a brief presentation of recently concluded project on distributed video coding called DISCOVER. This project is funded by European Union and adopts the Stanford approach.

2.3.1 California Wyner-Ziv Robust Video Coding Approach

UC Berkeley developed a video coding algorithm named PRISM [10], which takes advantage of DSC principles to allow flexible distribution of the computational complexity between an encoder and decoder and to achieve robustness to the drifts caused by channel loss. The algorithm aims to marry the intraframe coding features (low complexity encoding and robustness to transmission errors) with the interframe coding compression efficiency. The typical version of PRISM described in the literature [6, 9], is a block based algorithm in which the video frame is divided into *n-by-n* blocks, where a classification process is applied on each block. Each block is compared with the co-located block in the previous frame (zero-motion prediction) to find the *MSE*, the MSE of the difference between the two. By thresholding the (Minimum Mean squared) MSE, the block is then classified into one of many classes ranging from the SKIP class, where the block difference is so small; that it is skipped. The INTRA class, where there are so many differences that the block is encoded in the intra-frame mode. The blocks classified in the INTER coding classes constitute the major PRISM novelty. Therefore, the following sections will explain the encoding/decoding procedures of those blocks, in detail.

2.3.1.1 Inter Block Encoding Procedure

This block is encoded by syndrome coding to the lower frequencies DCT coefficients and the higher frequencies DCT coefficients are intracoded conventionally (see Figure 2-13). The encoder architecture is depicted in Figure 2-14; in the following its processes are described

a. Block wise DCT and zigzag scan: Block wise DCT is applied and the resulting DCT coefficients are zigzag scanned, based on the conventional MPEGx scanning order. The scanning starts from the DC coefficient and moves from lower frequency coefficients to higher frequency coefficients. In a real image, most of the energy is concentrated in a small number of DCT coefficients (lower frequency coefficients). Based on the energy concentration, typically the higher frequency coefficients have low or near-zero values, so no compression can be gained by compressing these coefficients. While the higher frequency is traditionally encoded the lower frequency coefficients are encoded, by using the syndrome coding.

b. Quantization The zigzag scanned DCT coefficients are then quantized to generate the quantized codewords. DC coefficients and a small number of AC coefficients are called Wyner-Ziv coefficients. The quantization step is determined by the correlation level of the block. The higher frequency coefficients are fed into other quantization module, typically the MPEG quantization scheme.

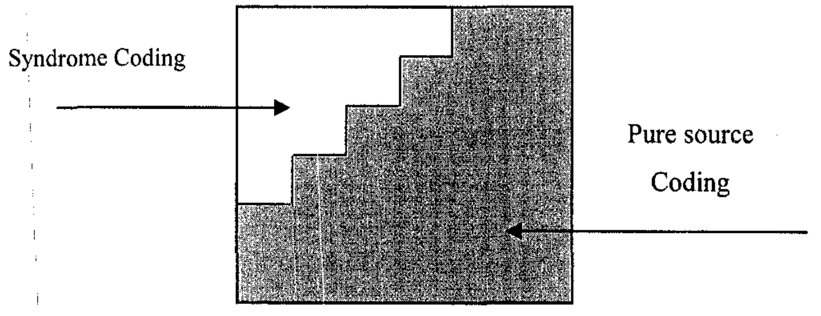

Figure 2-13: Selective encoding for the various transform coefficients within a block [40].

c. Syndrome Coding: the quantized codewords space is divided into several groups of codewords called cosets. Each cosets has an index label associated with it, this index is known as syndrome, and states the coset to which the codeword corresponding to the quantized transform coefficient belongs. The number of bits required to represent each coset is less than the bits needed to represent the codeword. In syndrome coding, only the coset index is sent instead of the codeword itself and therefore a compression is achieved. The resulting syndrome encoding bits are embedded within the bitstream syntax at the block level.

d. Refinement Quantization: In order to achieve a global desired reconstruction quality, a refinement of the base quantization step size is performed. This process corresponds to sub-partitioning the quantization interval in order to obtain a quantization step size corresponding to the desired quality. The sub-partitions within the base quantization interval are called refinement intervals; each has an index associated to it. The refinement interval index bits are also embedded in the bitstream syntax at the block level.

e. Entropy Coding: Another component of the bitstream syntax is the resulting bits of entropy encoding the quantized high frequency coefficients. These bits are called pure source coding bits.

f. Cyclic Redundancy Check: to help the decoder estimate the best predictor while performing the motion estimation task, the cyclic redundancy check (CRC) of the base quantized transform coefficients is computed and transmitted.

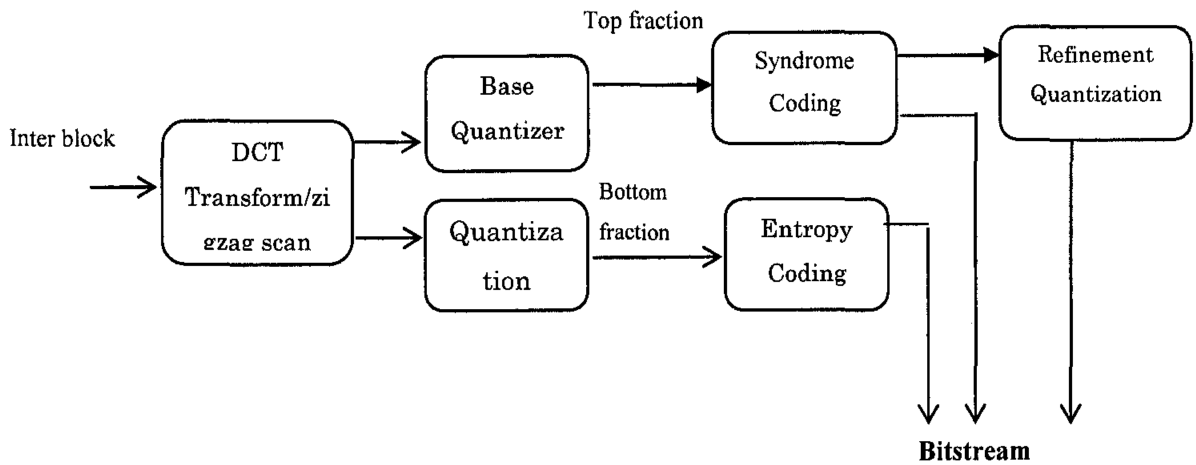

Figure 2-14: PRISM encoder architecture

At PRISM decoder side, the received syntax informs the decoder of the encoding mode of each block within the frame. The decoding of SKIP mode blocks is done by simply replacing the skipped blocks by the co-located blocks from the previous frame. The INTRA mode blocks are decoded conventionally (entropy decoding followed by de-quantization). The decoding of INTER mode blocks corresponding to the syndrome coding will be explained in detail since the PRISM novelty lies in the encoding/decoding of these blocks as we mentioned.

2.3.1.2 Inter Decoding of Syndrome Encoded Blocks

The processes associated with inter-decoded block is described in the following, the Figure 2-15 depicts the decoder architecture

a. Motion estimation: The task that is conventionally performed by the encoder is performed by the PRISM decoder to provide some information to help the decoder to decode the syndrome bits. This information consists of candidates "side information". The candidate predictors are obtained from previous frame by half-pel motion estimation.

b. Syndrome Decoding: to identify the right quantization index within the signaled coset. Different motion compensated predictors from previously reconstructed frames are tested in order to fill-in the missing most significant bits (MSB), The reason behind this scheme is that only those bits which cannot be obtained from the side information at the decoder (the LSB),i.e. The LSBs are actually the coset index. The best predictor matches the received CRC. The CRC is used as a reliable and unique signature for each block.

c. Base and Refinement De-quantization: The quantized coefficients are reconstructed the base quantization is inversed. A better quality can be obtained by using the refinement bits transmitted by the encoder.

d. The Reconstruction: Two estimates are obtained for the Wyner-Ziv coefficients: the first estimate is the coefficients obtained by syndrome decoding and inverse quantization to these decoded codewords. The second estimate is the coefficients of the best predictor found in the motion estimation stage. PRISM uses a linear estimation algorithm to estimate the final estimate of the Wyner-Ziv coefficients.

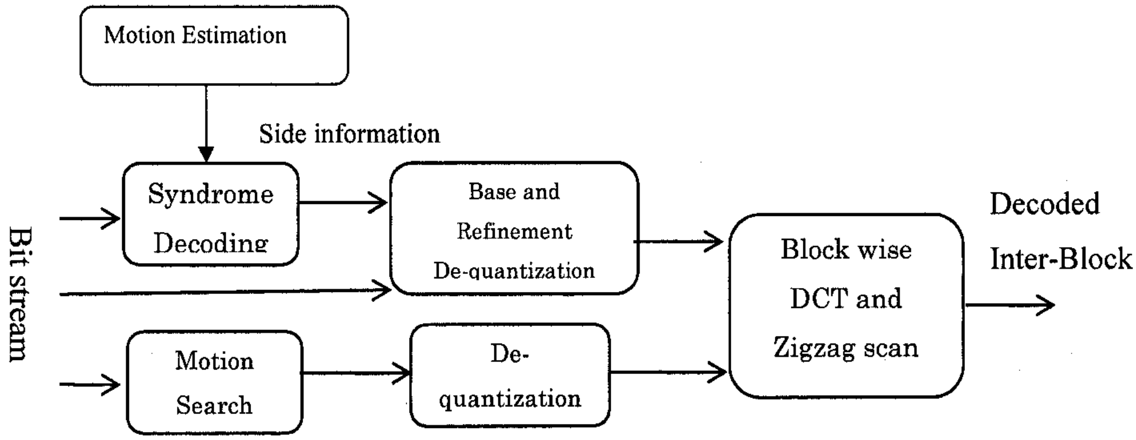

Figure 2-15: PRISM decoder architecture

2.3.1.3 PRISM Evaluation Experiments

In [6, 9], to evaluate PIRSM performance, the first 15 frames of Mother-Daughter, and Carphone videos sequences are used. The first of each sequence is intra-coded; the following frames are encoded by using the PRISM codec algorithm. The rate-distortion performance of PRISM is reported to be in between H.263+ intra and inter-modes, when the motion search is completely shifted at the decoder and a lossless channel is assumed (see Figure 2-19 & Figure 2-20).The experiments also have shown that the robustness of PRISM over H.263+ standard is much higher since annoying visual artifacts due to interframe error propagation are not observed. The loss of one frame when using PRISM has a negligible effect on the quality of the decoded video, since the error propagation is stopped because the PRISM encoder does not use the prediction loop as the H263+.

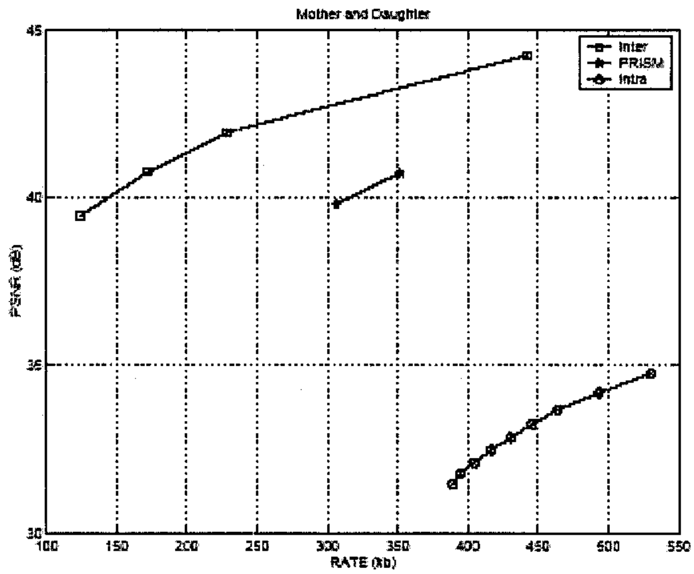

Figure 2-16: PSNR performance of the PRISM system for the Mother-Daughter sequence

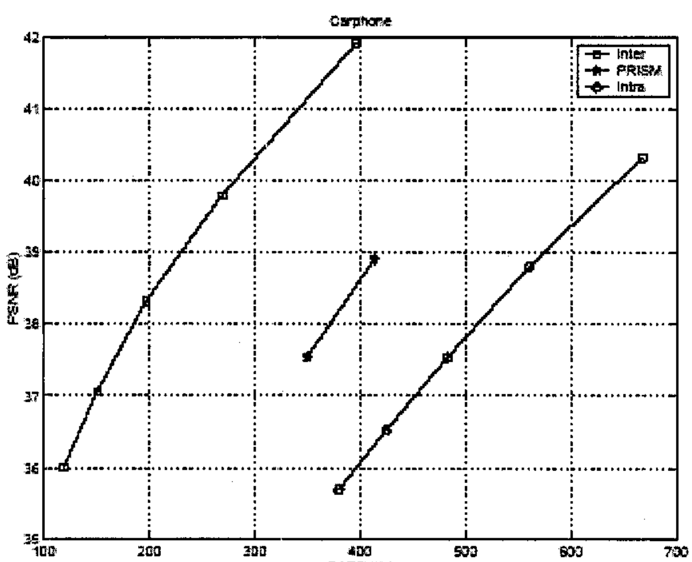

Figure 2-17: PSNR performance of the PRISM system for the Carphone sequence

2.3.1.4 PRISM Issue

The classification process mentioned in section 2.3.1.1, is typically performed by comparing the MSE of co-located block in two temporally adjacent frames, presents an undesired complexity in a supposedly low–complexity encoder. For the INTER classes, the encoder sends only the Least Significant Bits (LSB) of each quantized

DCT coefficient. Each of these classes has a predetermined number of LSB that need to be encoded for each DCT coefficient.

Offline training on several test sequences is used to obtain an estimate of these numbers. The correlation noise is being estimated at the considering the changing nature of the natural video some classes might end up receiving more or less LSB than they actually require. At the receiving end, the block is decoded by testing several predictors from previously decoded frames, until the decoded block matches the CRC sent by the encoder. The source statistics are unknown; therefore a successful decoding is not granted. In order to cope with the above drawback, work in [63] proposed to adaptively calculate the number of LSB for each coefficient, thus avoiding offline training. The limitation of this work is that two pieces of information need to be sent for each coefficient: the number of LSB and the actual value of the LSB.

2.3.2 Stanford Wyner-Ziv Robust Video Coding Solution

One of the first practical WZ video coding solutions has been developed at Stanford University by Bernd Girod's group [1-3, 11, 62]. It is also referred to as a feedback channel approach; this solution has become the most popular WZ video codec design in the literature, due to its simple encoder. There are several coding solutions proposed by this group. In all of these solutions the video sequence is divided in frames namely 'Wyner-Ziv' frames and 'Key' frames. The most recent solution proposed by the group is to send three types of bitstream to the decoder. One is the bitstream resulting from encoding Wyner-Ziv frames encoding, the second is the bitstream of the intracoded Key frames and the third is the supplementary bitstream, called the Hash code, which contains information about the current frame and it helps the decoder in the motion estimation. It has been implemented in the pixel domain (PD) where no transform is applied and in the transform domain (TD). In the pixel domain implementation, an intra-frame encoder and inter-frame decoder system is used [13]; where the decoder is responsible for exploiting all (or most of) the source statistics and, therefore, to achieving compression following the WZ coding scheme. A subset of frames from the sequence is designated as key frames. The key frames, K, are encoded and decoded using a conventional intra-frame codec. In between the key frames are Wyner-Ziv frames, WZ, which are intra-frame encoded but inter-frame decoded.

In the following we will present a more detailed description of the solution proposed by Bernd's group depicted in figure 2-18. The encoding/decoding of the key frame follows the conventional intra-frame codec using H263+. If the current frame to be encoded, the encoding process is performed using following five stages:

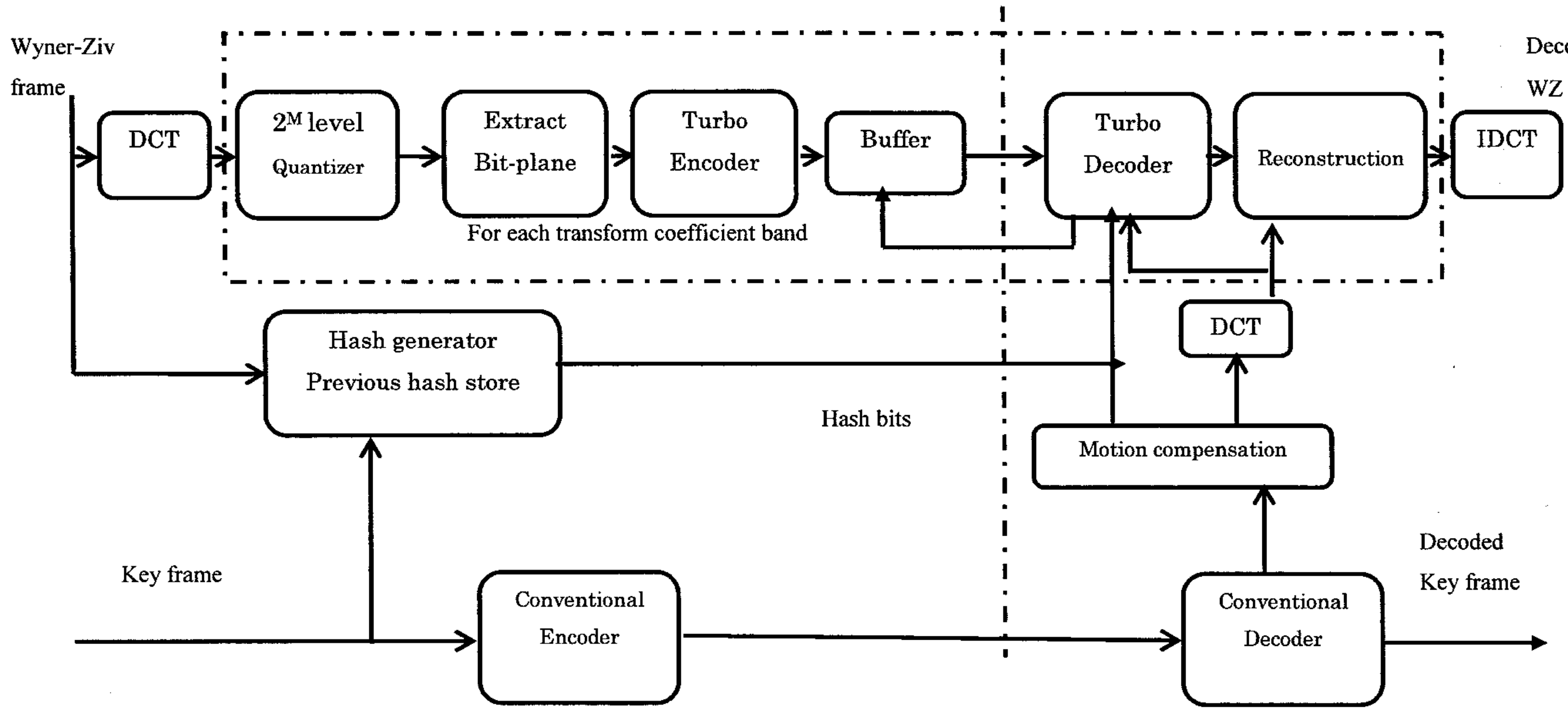

Figure 2-18: Stanford Wyner-Ziv video codec from [1]

2.3.2.1 Wyner-Ziv Encoding Procedure:-

The encoding process of the WZ frame is performed several steps as follows

a. Transform (DCT): The Wyner-Ziv frame is divided into *4-by-4* blocks; this step is followed by transforming the resulting blocks into discrete cosine transform DCT. The transformed coefficients of the entire frame are grouped together according to the position of the coefficient within the *4-by-4* blocks, forming transform coefficients bands.

b. Quantization: a uniform quantizer is designed for each transform coefficient band with 2^{K_t} levels. The quantization step is determined based on the number of levels. This stage results in the quantized symbol bands.

c. Extract bitplanes: the bits of quantized symbol band of the same index (importance, e.g. MSB) are grouped together forming bitplane arrays.

d. Turbo Encoder: The bitplanes arrays are fed individually to the Turbo encoder. The generated parity information for each bitplane is stored in the buffer and transmitted in small amounts upon decoder request via a feedback channel. The Slepian-Wolf codec is built based on Rate Compatible Punctured Turbo code structure.

e. Hash generator: Additional bits produced by the Wyner-Ziv encoder to help the decoder in the motion estimation task are used as a means to create the side information. The hash code corresponds to a small amount of W quantized transform coefficients; the hash codewords for the Wyner-Ziv frame are stored in a small hash memory in order to help in deciding whether a hash code must be transmitted to the co-located block in the next frame or not. The process of generating the hash code requires minimal computation and memory, to keep the complexity as minimum as the traditional intraframe encoding complexity.

2.3.2.2 The Wyner-Ziv Decoding Procedure

The key frames are decoded using conventional decoder and then used in the next adjacent Wyner-Ziv frame decoding to generate an estimate by means of motion estimation. The decoding of Wyner-Ziv is as following:

a. Motion-Compensation Extrapolation: The previously decoded key frames and the received hash bits are used to generate an estimate of the Wyner-Ziv frame. If no hash bits are received for the block, the corresponding block in the estimate of Wyner-Ziv is filled with the co-located samples from the previous decoded frame. DCT bands are then obtained with the same procedure was performed at the encoder side. To use the side information at the Turbo decoding stage, a statistical dependence model between corresponding coefficients bands is considered. The solution uses Laplacian distribution to model the difference between the corresponding bands in the original and the estimate of the Wyner-Ziv frame.

b. Turbo Decoder: For each transform coefficient band, the Turbo decoder starts decoding the most significant bitplane followed by the decoding of the remaining bitplanes. The received part parity bits and the side information together are used to decode the bitplanes, if the decoding fails the decoder requests more parity bits via the feedback channel; the usage of the feedback channel simplifies the rate control problem since the decoder knowing the available side information, can easily regulate the necessary bitrate. The request and decoding are repeated until the current bitplane error probability is lower than 10^{-3}; in this case the decoding is declared successful. In general, due to the availability of the side information, the number of Wyner-Ziv bits required to determine in which quantization interval (level) a transform coefficient is mapped to, from the 2^{K_t} possible levels, is lower than K_t bits and thus compression efficiency is achieved. By successfully decoding the last bitplane (LSB), the decoder has K_t bitplanes from which it can obtain the quantized symbol bands.

c. Reconstruction: the reconstruction of each transform coefficient band is performed based on the reconstruction function that uses the available quantized symbol and the side information.

d. Inverse discrete cosine transform (IDCT): A block based 4-by-4 inverse discrete cosine transform is performed on the reconstructed transform coefficients.

A simpler system in the pixel domain is also developed based on the Stanford solution, which generically makes use of a quantizer, a Turbo codec, a reconstruction module and an interpolation module.

2.3.2.3 Stanford Solution Evaluation Experiments

In order to evaluate the performance of the system two QCIF video sequences, *Salesman* and *Hall Monitor*, at 10 frames per second were used; Figure 2-19 and Figure 2-20 show the PSNR results obtained for the test content used. Each figure shows a set of curves corresponding to different GOP lengths: as it can be seen, the rate-distortion performance of the solution in [13] outperforms (up to 9 dB) the traditional DCT-based intraframe coding.

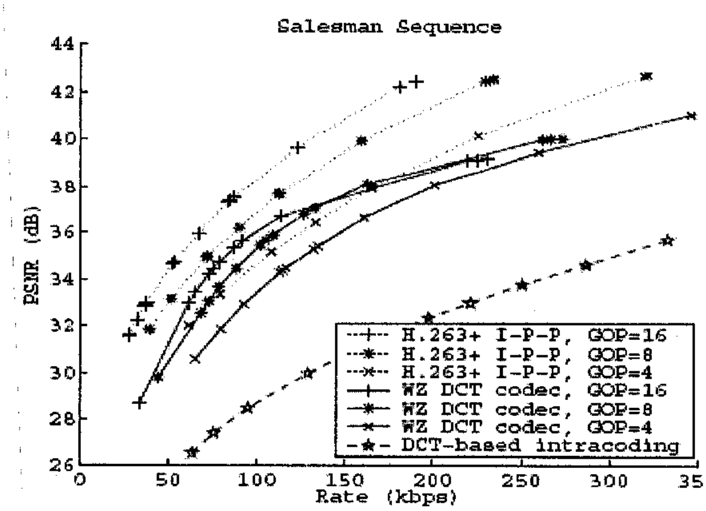

Figure 2-19: PSNR results for salesman [10]

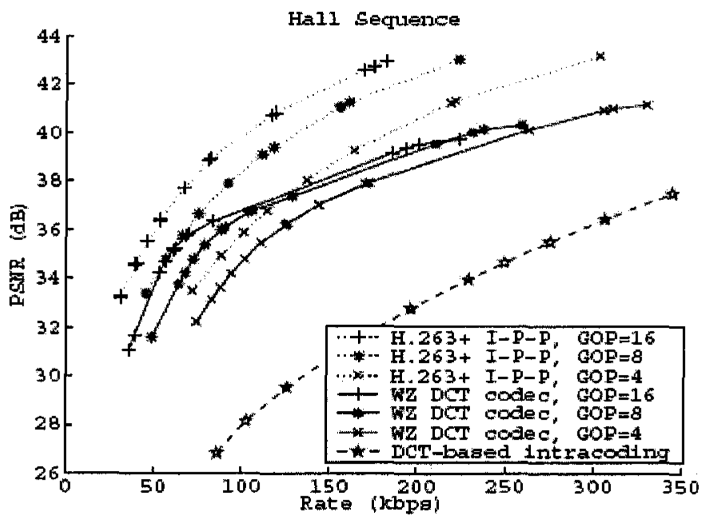

Figure 2-20: PSNR results for Hall Monitor sequence [10]

2.3.2.4 DISCOVER

A European project named DISCOVER, was funded under the European Commission IST FP6 program with the objective of exploring and proposing new video coding schemes. The area distributed video coding was expected to be for new emerging application such as video surveillance. It targets new advances in coding efficiency, error resilience, scalability and thus paving the way for a breakthrough video coding [64]. The project proposal adopts Stanford architecture and adds a new stage at the encoder. However, it differs from Stanford architecture, in that it uses different Slepian-Wolf codec and side information refinement algorithm. The details of these different modules are given here under:

- DISCOVER uses the rate-compatible LDPC Accumulate (LDPCA) codes as the Slepian-Wolf codec. An LDPCA encoder is composed from an LDPC syndrome-former concatenated with an accumulator. For each bitplane the created syndrome bits are accumulated modulo 2 to produce the accumulated syndrome. To establish if decoding is successful the convergence is tested by computing the syndrome check error, i.e. the Hamming distance between the received syndrome and the one generated using the decoded bitplane, followed by a cyclic redundancy check (CRC). If the Hamming distance is different from zero, then the decoder proceeds to the next iteration. After a certain amount of iterations (experimentally it was found that 100 is enough), if the Hamming distance remains different from zero, then the bit plane is assumed to be erroneously decoded and the LDPCA decoder requests for more accumulated syndromes via the return channel.

- In motion compensation interpolation (MCI) module, a motion field closer to the true motion is estimated between backward X_b (past) and forward X_f (future) reference frames; then motion compensation between the two references is performed to obtain the side information. A block matching based on a modified mean absolute difference (MAD) criterion is used in order to regularize the motion vector field, which favors motion vectors closer to the origin. Then, bidirectional motion estimation is performed in order to find symmetric motion vectors from the current WZ frame to P_B and P_F. Spatial motion smoothing based on a weighted vector median filter is applied afterwards to the obtained motion field to remove outliers. Finally, motion compensation is performed between P_B and P_F along the obtained motion field, so as to generate the side information.

- Minimum rate estimation module in order to reduce the number of accumulated syndrome requests to be made by the decoder (which has a strong impact on decoding complexity), the encoder estimates a minimum number of accumulated syndromes to be sent per bitplane and per band.

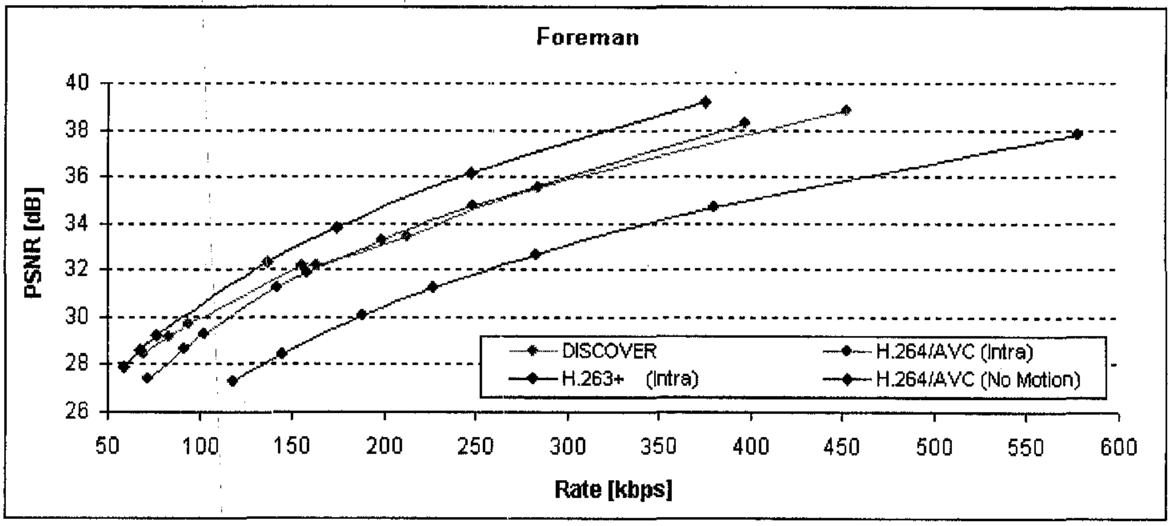

Figure 2-21: PSNR results for Foreman [63]

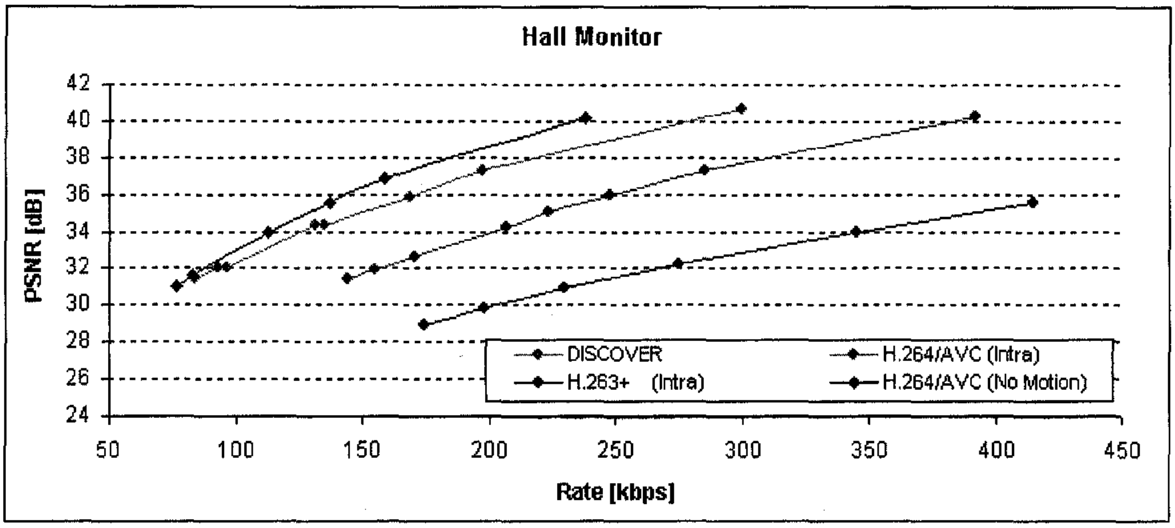

Figure 2-22: PSNR results for Hall Monitor [63]

The following tests conditions have been used to obtain the example rate-distortion (RD) results are presented here: All frames of "Hall & Monitor" and "Foreman" sequences are QCIF@15Hz with a GOP length of 2. For the motion interpolation, ±32 pixels are used for the search range of the forward motion estimation; both references are first low pass filtered with a 3×3 size mean filter. Key frames are always encoded with H.264/AVC Intra (Main profile), and the quantization parameters (QP) for each RD point are chosen so that the average quality (PSNR) of the WZ frames is similar to the quality of the key frames. All rate and distortion results refer only to the luminance. The experimental results in Figure 2-21and Figure 2-22 show that the presented codec is already RD competitive when compared with other codecs with similar (low) encoding complexity, even though there is still more room for research.

2.3.2.5 Design Issues in Frame Based DVC

This section highlights some design issues that the frame based system, presented in figure 2-18 is still plagued with, namely, the poor correlation noise model, the poor rate control and the non-adaptive quantization.

Issue 1 – Inaccurate Correlation Noise Model: Although more sophisticated motion estimation algorithms can be used to generate the side information, still the practical frame prediction is fundamentally faulty due to events like occlusions. Occlusions create noise in predicted frames with two important properties. Firstly, this noise is very location-specific and it is always located at the edge of a moving object or the frame edge in the case of camera motion; Secondly, occlusion noise is hard to characterize [16]. In regions where the motion estimation completely fails, caused by occlusion, the side information is uncorrelated with the original data. The noise process then models the effect of occlusions and will have a large variance and follows, for instance, a uniform PDF model. In the non-occluded regions the noise process is assumed to be well-behaved, and follows, for instance, a Gaussian or Laplacian PDF model with a small variance [16]. Distributed source coding relies heavily on efficient error correcting codes; the performance of these codes depends greatly on the choice of the noise model that characterizes the virtual dependency channel [15, 54]. The lack of information about the unpredictable occlusion noise can be tackled by the use of interleavers in the Turbo based systems; interleaving eliminates the location specificity of the noise, with cost of rate penalty.

Issue 2 – Non-adaptive Quantization: The frame based DVC solution assumes that the quality of the side information is constant across the whole frame. Therefore it uses one quantizer for the entire frame. Considering the varying motion characteristics and the amount of detail, the estimation of the original frame at the decoder is not the same for all regions of the frame. Therefore, to achieve certain operational characteristics in terms of distortion reduction the quantizer must consider the quality of the side information.

Issue 3 – Poor Rate Control: The feedback channel (FC) has the role to adapt the bit rate to the changing statistics between the side information and the frame to be encoded. The frame based DVC solutions assume that these changing statistics are constant across the whole frame. Each pixel of the WZ frame is scanned row-by-row and then uniformly quantized. The quantized symbols form an input block which is binaried and fed into a Turbo encoder for coding and transmission. Therefore, the size of input block is assumed arbitrarily large because it represents all bits from the whole frame or one bitplane at once [11, 65]. Too large an input block will produce significant computation latency during the encoding and decoding process [66]. The system will not be able to provide a timely WZ decoded output due to the vast Slepian-Wolf turbo coding and decoding delay [66]. Frame based DVC systems make use of a rate-compatible-punctured codes (RCPT) codec as the module of the SW codec. The encoder "blindly" sends the parity bits according to the puncturing pattern determined by the next lower coding rate. This causes a waste of transmission when some unnecessary parity bits are sent to help decoding some bits which are already correctly decoded [2].

Issue 4 – Applicable only in bi-directional scenarios: A typical weakness of this architecture is the usage of the FC channel itself. The usage of a feedback channel is not realistic for applications that are unidirectional and therefore it is more than a weakness in that context.

2.3.2.6 Related Literature

In recent years, a significant number of research groups around the world have adopted the Stanford WZ coding architecture and changed it in order to cope with the above issues. This subsection will briefly describe some examples of work that address these issues individually table 2-1 briefly describe the methods proposed in these work examples.

In [57], in order to exploit the spatial variability of the correlation noise, the parameter of the model is calculated at the block level; the considered block size is equal to the one used in the frame interpolation process (8-by- 8) in order to more easily match the frame interpolation errors; the resulting RD performance is outperformed the frame level based noise channel modeling. A non-stationary model to characterize the correlation noise is proposed in [67] where two models are used; one to model the non-occluded regions and the other to model the occluded regions, the results show that the performance of the Soft-input-Soft-output (SISO) decoder, and therefore of the overall system, can be improved greatly by classifying the decoder-generated side-information into two (or more) reliability classes. It also shows that the channel model should be an accurate representation of the real behavior of the channel; otherwise the decoding performance will heavily degrade. The lengthy block problem has also been addressed in [66, 68] by dividing the WZ frame X into M sub-images X_m, $m = \{1,2, \dots . M\}$ and each sub-image is independently encoded using a Turbo-code based Wyner-Ziv encoder. Therefore, the block size of the Turbo encoder decreases to $1/M$ of the frame size. In the proposed system [68], Block-Adaptive Wyner-Ziv coding BAWZC-based DVC system, the puncturing rate is independently adjusted for each block and the BAWZC decoder has to perform by informing the WZC encoder which blocks needs more parity bits. The system proposed in [66] restricts the length of the SW input to reduce the decoding complexity and the resulting delay. Both works [66, 68] assume the stationary noise and use the same correlation noise model to characterize the noise over all the sub-images; which makes no point to the sub-images other than just reducing the lengthy block; they alleviate the problem of spreading the estimation errors all over the bitstream, by keeping the localized errors close to each other in the generated symbol stream, in the results for both [66, 68], it is observed that system performance is further enhanced.

The first attempt to solve the feedback problem was proposed in [69]where the encoder averages the previous and the next frame, and the difference with the original frame is calculated and the error probability computed. The required puncturing level for the Turbo encoder is then determined from this probability using an algorithm based on empirical results obtained by examining three different standard test sequences. An algorithm targeting free FC distributed video coding system is proposed in [56]. The algorithm requires generating low quality version of the side information Y_{enc}. A simple frame interpolation was also proposed in [56], where first the $H(X|Y_{enc})$ is computed. Extra corrective rate (p) is added due to the difference in side information version at the decoder Y_{dec} In [70] another work is proposed to remove the feedback channel. First an error probability P_e is estimated then an adequate number of parity bit sets that are sufficient to decode with a residual error probability $\mathcal{A}_{ek}$ below a threshold t_k $(\mathcal{A}_{ek} < t_k)$ are estimated. Suitable cost functions are estimated to provide the residual error probability $\mathcal{A}_{ek}$ as a function of P_e and the number of parity bit sets. These functions are estimated by averaging simulations over a set of input random sequences with different error probabilities and then used to calculate bit-rate such that the residual error probabilities are lower than the threshold t_k. A rate control algorithm for pixel-domain Wyner-Ziv video coders is presented in [71]. The algorithm assumes three different modes to encode the considered frame (INTRA, SKIP, WZ). The rate control selects the mode (SKIP, WZ and INTRA) of each WZ-frame and the quantization step size of all the frames. In

[72] feedback free DVC architecture is presented with focus on eliminating the dependence on the feedback channel and the ideal correction detection at the decoder based on original frames. The encoder rate control is formed by a residual estimation and a decision tree. The Machine Learning tools are used to get the decision tree for the statistics extracted from the residual frame. A decision tree is made by mapping the observations about a set of data in a tree made of arcs and nodes. For each 4×4 block, the proposed algorithm uses the mean and the variance of 2×2 sub-blocks, the 4×4 variance and the 4×4 mean. The corresponding number of requests for this block is also used in the function as determined by the ideal feedback based architecture reference [2], using the same block size. A data mining tool is used to discover a pattern of residual information for the number of request over a feedback based architecture. The training set was made with the first frame of some sequences in order to capture the different aspects which can appear in the video.

Table 2-1 Related Literature

Ref.	**Year**	**Method**	**Merits**	**De-merits**
[57]	2008	Correlation noise modeling at block level: to obtain accurate channel model	- Location-specific channel model	-Non-adaptive source coding. - Poor rate control. - Lengthy SW input.
[67]	2005	Non-stationary model: tow models for occluded and non-occluded regions	Location-specific channel model	-Non-adaptive source coding. - Poor rate control. - Lengthy SW input.
[68]	2007	Block-Adaptive Wyner-Ziv coding BAWZC: the puncturing rate is independently adjusted for each block	- Efficient rate control	-Non-adaptive source coding. - Stationary channel model. - Lengthy SW input.
[66]	2009	Restrict SW input length: to reduce the complexity of the Slepian-Wolf codec	- Short SW input. - Efficient rate control	- Non-adaptive source coding. - Stationary channel model.

As summarized in table 2-1, the literature is replete with many attempts that try to address the performance issues of Frame-Based DVC codec. Each attempt, however, has addressed one or the two issues at a time, and not the need for adaptive source coding which reflects efficient rate allocations considering the spatial variation of the video contents. Table 2-1 also shows that these attempts have failed in developing a solution that collectively and simultaneously addresses those issues. Hence, one can conclude that the design of an optimal DVC codec is still an open research problem - a design that addresses all these issues collectively while adhering to the DVC constraints.

2.4 Chapter summary

This chapter presented the theoretical foundation of the DSC principle and the two major information-theoretic results that laid the bases for developing practical DVC codec solutions. The chapter also explained how the source coding, channel coding and estimation interplay to construct a DVC codec solution. Some of the architectural developments which are derived from the basic architecture were given. The chapter presented the works of the two research groups who have been responsible for the development of the most relevant distributed video coding systems available: Bernd Girod's group at Stanford (University of Stanford) and Kannan Ramchandran's group at Berkeley (University of California). The chapter concluded by highlighting some of the issues in current DVC approaches and their related works in the literature. Extra attention is given to Stanford approach since it is the most popular DVC codec solution due to its low encoding complexity. The related works concluded in this Chapter serves as the motivation of the proposed solution.

CHAPTER 3

REGION-BASED ADAPTIVE DVC WITH FEEDDBACK CHANNEL

Present DVC systems tackle the problem of distributed video coding at two different level of granularity, namely, block based systems (PRISM) [6, 9, 10] and frame based systems [11, 13, 65]. In this work a different granularity level is proposed that addresses coding issues of presents DVC systems collectively. The video coding application scenarios can be divided into unidirectional scenario and bidirectional scenario. The current frame based DVC systems use the feedback channel to perform the rate control. This chapter is dedicated to introducing the proposed Region Based Adaptive DVC codec with feedback channel solution. Next chapter however, focuses on feedback free DVC codec solution. Section 3.1, introduces the proposed DVC system and explains the objectives that the proposed system is meant to achieve. Subsection 3.1.2 gives a coarse representation to the proposed solution. Subsection 3.2.1 gives a simplified walkthrough of the proposed Region-Based Adaptive DVC codec in Pixel domain. In section 3.2.2 a simplified walkthrough of the proposed Region-Based Adaptive DVC codec in Transform domain. The detail of implementing the proposed solution in Region-Based Adaptive DVC codec pixel domain is presented in section 3.3. The detail of implementing the proposed solution in Region-Based Adaptive DVC codec in transform domain is presented in subsection 3.4. Summary to the chapter contents is presented in section 3.5.

3.1 Proposed Region-Based Adaptive DVC Technique

3.1.1 Background

In traditional video coding, the first frame of a group of frames is sent as it is and the subsequent frames are compressed by performing motion estimation/compensation to exploit the temporal redundancy. The displacement of blocks between successive frames is first estimated, task known as (motion estimation). The resulting motion information that represents extra information to be coded is then exploited in efficient interframe predictive coding. Both the motion information and prediction error are sent to the decoder, where the prediction error is added to the motion-predicted frame. Further compression is achieved through the exploitation of the spatial redundancy in motion fields associated with each block to avoid transmitting motion vectors multiple times. A region-based approach was proposed for traditional video coding systems in [73-76], where the region is a collection of blocks that have the similar motion model (translational, affine, etc) or have similar rate-distortion characteristics. In particular, region-based compression methods describe the images in terms of a set of regions, each unit with a unique coding unit coded and decoded separately The rate-distortion information is estimated based on the local characteristics of the region.

In region-Based traditional video coding the frame partitioning is performed by considering the current frame and its predicted version. In typical DVC solution, the predicted frame is obtained by an estimation process using two key frames. This frame is called "side information" which is only available at the decoder. Since the current frame is not available at the decoder, another estimate has to be employed to obtain the current frame. In addition, the low-complexity encoding constraints hinder employing traditional Region-Based solution as it is. Since information refereeing frame partitioning is unavailable at the encoder, implementing the Region-Based video coding scheme represents a real challenge in the DVC setting.

3.1.2 Region-Based Adaptive DVC Codec Higher Level Abstraction

In conventional video coding the Region-Based video coding can benefit the DVC system by addressing the frame-based system issues (section 2.3.2.5) collectively. If employed successfully the Region-Based video coding scheme in DVC is expected to offer the following advantages over other DVC scheme.

1. Overcome the problem of inaccurate dependency channel model and incorporate a location-specific dependency channel model. The location-specificity encourages

on considering local characteristics of the WZ frame. For example, it is known that the dependency channel model of the background area is different from that of the foreground area that contains motion information. Many a times, the foreground itself may be a superposition of more than one motion pattern and accordingly may need to be decomposed into its constituent regions representing separate classes of motion. State-of-art-the-art Frame-Based DVC codec solutions are not able to represent the WZ frame as accurately as Region-Based segmented frames.

2. Improve the rate control process: since, by applying a Region Based DVC solution, the encoder "blindly" sends the parity bits to the locations of decoding errors which are limited into one region rather than a whole bitplane. Thus the parity bits sent are more accurate. Hence, a different level of protection is allocated for each region, for instance, the high motion area causing failure of decoding only belongs to one region. The decoder requests more parity bits only when this region is sent. Before that, for other regions, like region only containing the background information, the decoding will be successful and no requests for more parity bits will be made.

3. Adaptively allocate a rate based on the local video characteristics: Again unlike other DVC solution, the Region-Based scheme will partition the WZ frame into separate coding units, each unit with it own rate-distortion characteristics. The scheme makes use of that and maximizes the quality of the reconstructed WZ frame while minimizing the overall rate. Region structure allows applying appropriate source coding for each region targeting a certain operational RD point. The reconstruction distortion can be further reduced if a quantizer is selected based on the initial estimation distortion of the region.

Accordingly, this work proposes a Region-Based Adaptive DVC solution that borrows the Region-Based concept of traditional video coding and employs novel implementation to achieve an efficient and novel DVC solution. The solution proposes partitioning the WZ frame into many coding units, each unit having different DVC parameters. The scheme of partitioning a frame into regions can be seen as allocating a different channel and rate for each region. Figure 3-1 shows an abstract level representation of the proposed solution components. In the following subsections the function performed by each of the component is briefly described.

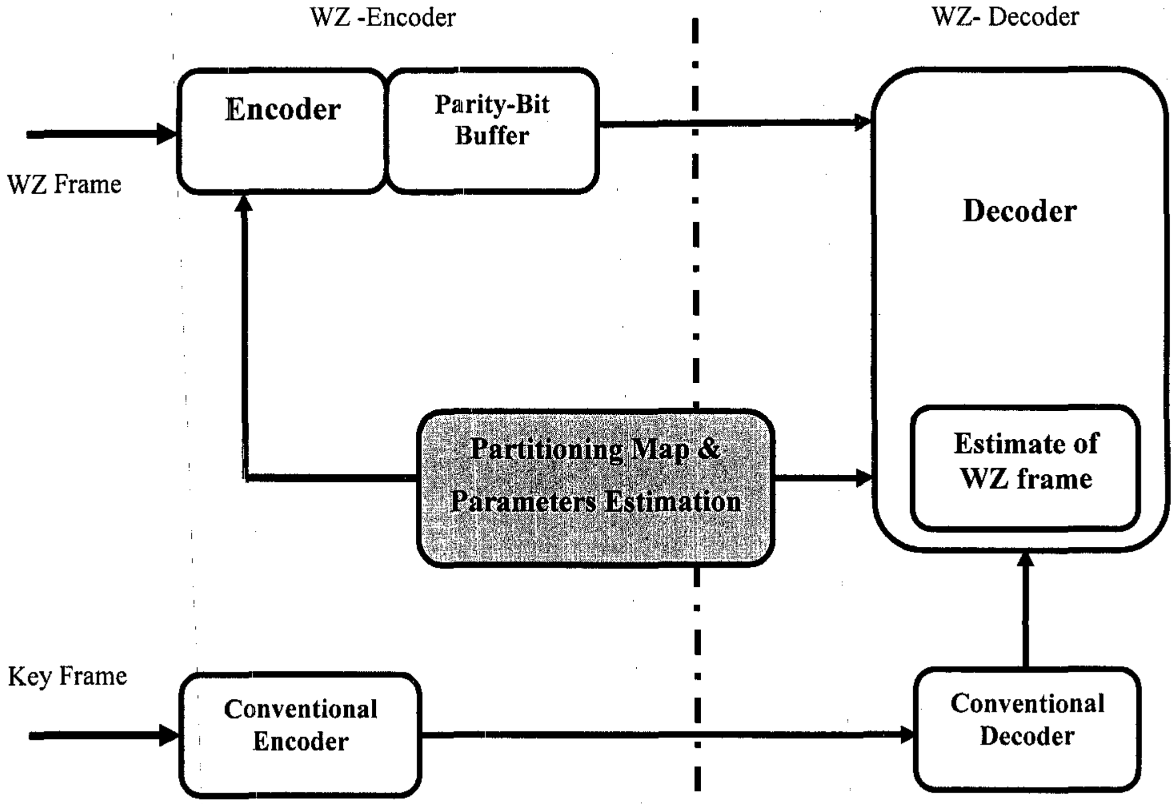

Figure 3-1 Higher Level Abstraction Architecture

3.1.2.1 The Frame Partitioning and DVC Parameters

The frame partitioning is meant to divide the original WZ frame "side information" into several partitions based on the local characteristic of the video sequences. That allows DVC parameters for each region to be allocated adaptively based on thier local characteristics. The estimation of the DVC parameters (dependency channel, quantizer and rate) at a finer level is expected to enhance the overall performance of the DVC solution. In DVC literature, the location of where the partitioning process and parameters estimation is carried out depends on the targeted solution as shown Figure 3-1. It can be performed either at the decoder or at the encoder. The partitioning map then is communicated via the feedback/forward channel to the encoder/decoder to generate the corresponding partitions. The region map must also be compressed and communicated to the encoder/decoder, so each region is decoded with the aid of its corresponding parity bits generated by the encoder and its corresponding region at the decoder estimate "side information". To maintain the overall simplicity of the DVC encoding part the partitioning and the DVC parameters estimation process is implemented in a way that it reutilizes the existing system components and adds little extra complexity to the existing components. In each one of the two cases there is a challenge of how to perform the partitioning. In case of performing the partitioning at the encoder the challenge appears in generating the side information at the encoder while preserving the encoder simplicity. While performing the partitioning at the decoder is challenged by the absence of the original WZ frame.

3.1.2.2 WZ-Encoder

Typical WZ-encoder performs two coding tasks, the source coding (quantization) and the SW codec. Since the partitioning process results in different coding units each unit is encoded separately at the WZ-encoder. The WZ-encode is fed by the coding unit/region to perform the source coding process resulting quantized symbols, which are binariesed and fed into a SW encoding. The SW encoding performs the channel coding to generate the parity bits for each bitplane, which divided into packets and stored in a buffer. To achieve compression only an appropriate number of the stored packet is transmitted to the WZ-decoder. This number of packet is decided by the rate control process either carried out by the decoder via feedback channel.

3.1.2.3 WZ-Decoder

In a typical DVC system the WZ-decoder performs most of the compression process, for the most complex part (exploiting the temporal redundancy) of the compression is shifted to the decoder. The WZ-decoder encloses three modules, namely, the side information generation (estimation of current WZ frame), SW decoding and reconstruction. The received parity bits packets and the side information are fed into the SW decoding resulting in the quantized symbol. The compressed region is reconstructed at the reconstruction by using the side information and quantized symbol.

3.1.3 Proposed Region-Based Adaptive DVC Codec Features

3.1.3.1 ***Feature 1:*** *Accurate Dependency Channel Model*

Slepian-Wolf coding relies heavily on efficient error correcting codes. The performance of these codes depends greatly on the choice of the noise model that characterizes the virtual dependency channel [54]. In other words, the performance is improved if the estimation error model is accurate. In [67], it was concluded that the behavior of the virtual dependency channel is substantially more complicated than a simple binary symmetric channel (BSC) or additive white Gaussian noise (AWGN) channel model often assumed in communication systems that use channel codes. The complicated nature of such signal is best described by a mixture of model rather than simple homogenous model. To validate the hypothesis that the mixture models best fit the dependency channel, an experiment is conducted by using the Kolmogorov-Smirnov goodness-of-fit test. Goodness-of-fit tests are used to examine a hypothesis that a given data set comes from a model distribution with given parameters. The Kolmogorov-Smirnov (KS) test is a popular goodness-of-fit test. The KS statistic test measure the distance between the empirical Cumulative Distribution Function (CDF) and the model CDF which measures the goodness-of-fit. If the empirical CDF is tested against several model CDFs, the model that gives the minimum KS statistic (distance between the empirical CDF and the model CDF) can be taken to be the best fit for the data. Kolmogorov-Smirnov values were obtained and presented in Table 3-1, for different video sequences. For a typical significance level of 5%, those values imply accepting the hypothesis that it follows a Laplacian mixture model. The minimum statistics and the best fit distribution for a dependency channel have been indicated in bold.

It is not surprising that the mixture model fits better the dependency channel for it considers the non-homogenous nature of the dependency channel as utilized by traditional video coding. From the table it can be seen the best fit mixture is the Laplacian mixture, a mixture of Laplacian and Gaussian has not been tested. This indicates that estimated WZ frame consists of different region that have more estimation error in one region than the other.

Table 3-1: KS goodness-of-fit test

Video sequence	Resolution	Laplacian	Gaussian	Lap Mixture	Gauss Mixture
Carphone	144X176	0.3533	0.5423	**0.0248**	0.0477
Miss America	144X176	0.2504	0.7409	**0.0439**	0.0644
Foreman	144X176	0.3011	0.6840	**0.0348**	0.0641
Hall Monitor	144X176	0.3435	0.5856	**0.0335**	0.0428
Coast Guard	144X176	0.3257	0.5645	**0.0218**	0.0427
Akiyo	144X176	0.3350	0.6340	**0.0391**	0.0656

3.1.3.2 ***Feature 2:*** *Appropriate DVC Parameters Allocation*

The allocation of the appropriate DVC parameters (dependency channel, quantizer and rate) for each coding unit/region is adaptively done at the decoder. The dependency channel for the resulting partition/region is estimated at the decoder by using the corresponding partition in the two motion compensated key frames. The quantizer for each partition is selected at the decoder as well by considering the initial estimation error and the overall rate aiming to maximize the quality and minimize the rate. The rate (number of parity bits) for each region is allocated during the decoding by the decoder via the feedback channel.

3.2 Proposed Solution Lower Level Abstraction

The proposed partitioning concept is employed in designing DVC solution named Region-Based Adaptive DVC with feedback channel. The proposed solution is implemented in two domains as follows

1. Region-Based Adaptive DVC with feedback channel in Pixel Domain
2. Region-Based Adaptive DVC with feedback channel in Transform domain

3.2.1 Region Based Adaptive DVC with Feedback Channel in Pixel Domain Architecture

Region Based Adaptive DVC in pixel domain solution shown in figure 3-2 is the simplest DVC encoder where the compression is achieved by exploiting only the temporal redundancy at the decoder. The architecture of the proposed Region Based Adaptive DVC with feedback channel in pixel domain as shown in figure 3-2 can be seen as a finer representation of the higher level abstraction architecture in Figure 3-1.

The three main components of the higher level abstraction architecture are detailed as following:

Encoder

–Conventional coding for key frames.

–Source coding (Scalar Quantizer) for WZ frame.

–Slepian-Wolf Coding for WZ frame.

Frame partitioning and DVC parameter estimation

- Frame partitioning and DVC parameters estimation (dependency channel, Quantizer)

Decoder

–Conventional decoding for key frames.
–Frame partitioning

–DVC parameter estimation.

–Side information generation

–Slepian-Wolf decoding.

–Reconstruction

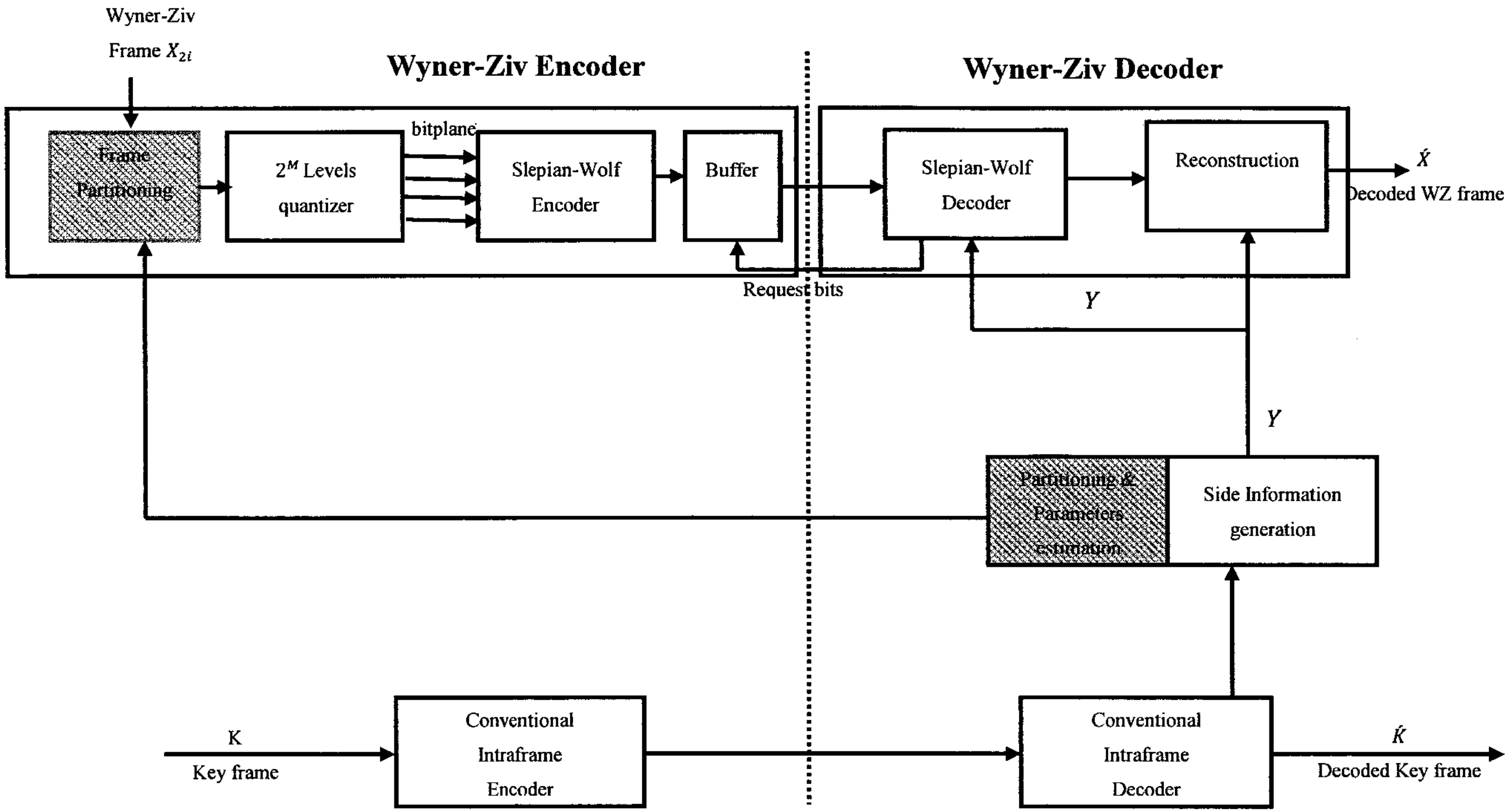

Figure 3-2: Architecture of the Region-Based Adaptive DVC codec in pixel domain

In the following a simplified Walkthrough the scheme is presented

1. The key frames (odd frames) are conventionally encoded and sent to the decoder (conventional encoding/decoding).
2. The side information process performs the motion estimation and estimates the forward/backward motion vectors at the decoder.
3. The two motion-compensated key frames are produced.
4. The partitioning process is performed at the decoder and the side information is partitioned into L regions $\{Y_1, Y_2, \ldots Y_L\}$ the corresponding regions map is created by using quad tree algorithm (frame partitioning).
5. The DVC coding parameters (Quantizer, dependency channel model) for each region are obtained (DVC parameters estimation).
6. The region map and the DVC coding parameters are transmitted to the encoder.
7. The encoder decomposes the compressed regions map and partitions the original frame into corresponding L regions $\{X_1, X_2, \ldots X_L\}$ (frame partitioning).
8. The samples of the first region X_l are quantized, generating the quantized symbol stream (Scalar Quantizer).
9. The Slepian-Wolf encoder produces redundant information (e.g. parity bits) from the quantized symbol stream (Slepian-Wolf encoding).
10. At the decoder, the quantized symbol stream is decoded through joint source-channel decoding with the aid of the region X_l , the side information (Slepian-Wolf decoding).
11. The decoded quantized symbol stream and the side information Y_l are then used together in a reconstruction module to reconstruct the X_l sequence $\acute{X}_l$. The region Y_l (typically a decoder estimate of the region X_l to be Wyner-Ziv encoded) is considered to be available at the decoder (Reconstruction).
12. The system repeats steps 6-9 for all the remaining regions.

3.2.2 Region-Based Adaptive with Feedback Channel in Transform Domain

Region Based Adaptive DVC codec in transform domain shown in figure 3-3 is more complex than the pixel domain implementation due to the DCT, which helps exploit the spatial correlation among the region contents. The architecture of the proposed Region Based Adaptive DVC with feedback channel in transform domain as shown in figure 3-3 can be seen as a finer representation to the higher level abstraction architecture in Figure 3-1. The three main components of the higher level abstraction architecture are detailed as following

Encoder:

–Conventional coding.

–quantizer)

–DCT transform.

–Source coding (Dead-Zone Quantizer).

–Slepian-Wolf Coding.

Frame partitioning and DVC parameter estimation

- Frame partitioning and DVC parameters estimation (dependency channel, Dead-Zone Quantizer)

Decoder

–Conventional decoding

–Frame partition.

–DVC parameter estimation.

–Side information generation

–Slepian-Wolf decoding.

–Reconstruction.

–Inverse the DCT transform.

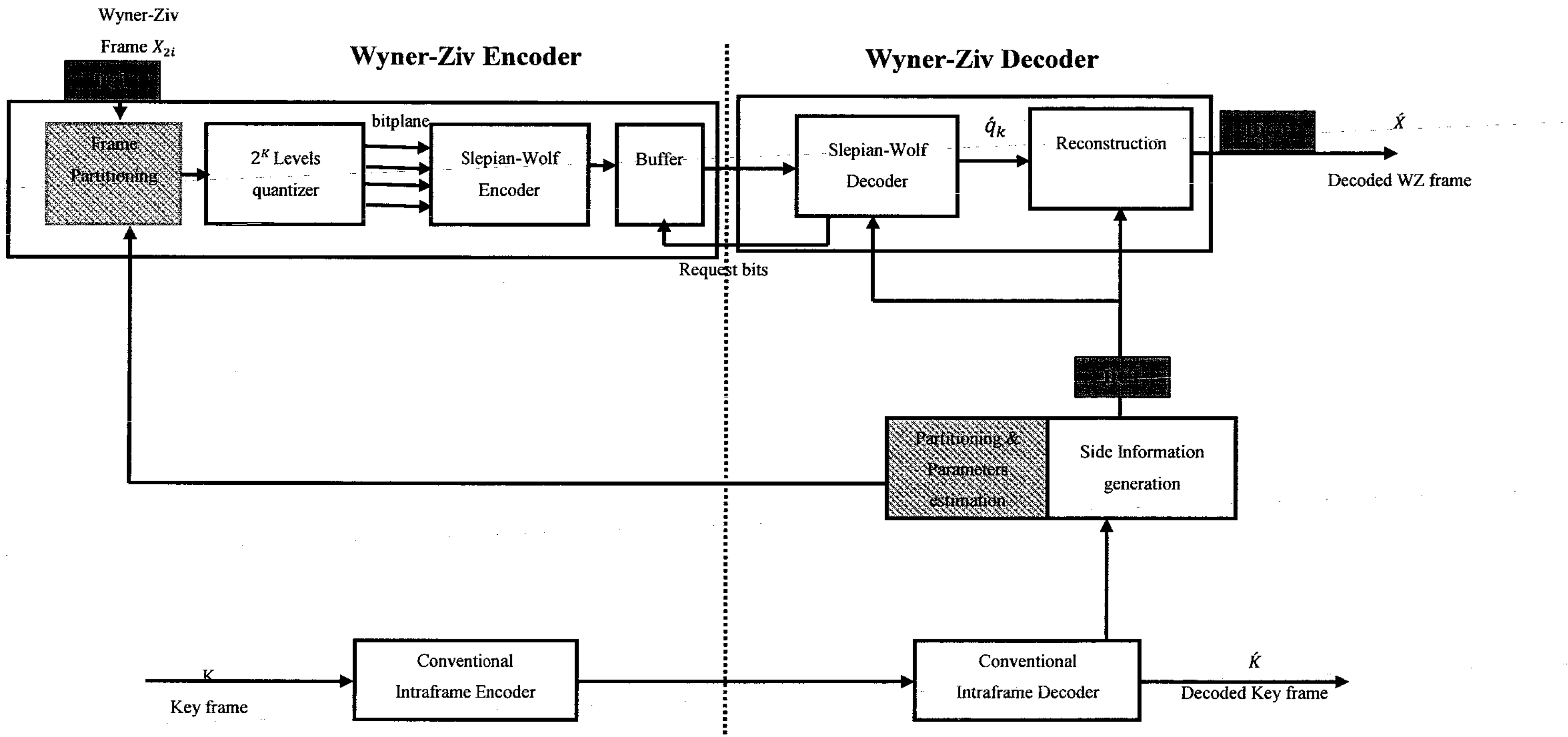

Figure 3-3 Architecture of the Region-Based Adaptive DVC codec in Transform domain

A simplified walkthrough as follows

1. The key frames (odd frames) are conventionally encoded and sent to the decoder (Conventional encoding/decoding).
2. The side information process performs the motion estimation and estimates the forward/backward motion vectors at the decoder.
3. The two motion-compensated key frames are produced.
4. The partitioning process is performed and the side information is partitioned into L regions $\{Y_1, Y_2, \ldots Y_L\}$ the corresponding regions map is created by using quad tree algorithm.
5. The DVC coding parameters (Quantizer, dependency channel model) for each region DCT bands are obtained.
6. The regions map and coding parameters are transmitted to the encoder.
7. The encoder decomposes the compressed regions map and partitions the original frame into corresponding L regions $\{X_1, X_2, \ldots X_L\}$ (Frame Partitioning).
8. The first region X_l is transformed into DCT transform (DCT transform).
9. The DCT bands of the first region are formed and then quantized, generating the quantized symbol stream (Quantization).
10. The Slepian-Wolf encoder produces redundant information (e.g. parity bits) from the quantized symbol stream (Slepian-Wolf encoding).
11. At the decoder, the quantized symbol stream is decoded through joint source-channel decoding with the aid of the region Y_l, the side information (Slepian-Wolf decoding).
12. The decoded quantized symbol stream and the side information Y_l are then used together in a reconstruction module to reconstruct the X_l sequence, $\acute{X}_l$ Reconstruction*)*.
13. Inverse DCT.
14. The system repeats steps 6-10 for all the remaining regions.

3.3 Region-Based DVC codec In Pixel Domain - Theoretical Background

Figure 3-2 depicts the proposed system architecture. In this section the implementation and the theoretical basis of the proposed Region-Based Adaptive codec DVC solution in pixel domain is presented. This system in the light of the higher level abstraction architecture contains the following parts

- The Frame partitioning and DVC parameters Estimation.
- The Encoder: Source coding and SW encoding.
- The Decoder: Side information generation, SW decoding and reconstruction.

The following subsections will explain these parts in detail.

3.3.1 Frame Partitioning and DVC Parameters Estimation

This part represents the new module added to the Frame-Based DVC codec system. In this module the WZ frame is partitioned into several coding units and the appropriate DVC parameters are estimated. Since the side information is unavailable at the encoder, this task cannot be performed at the encoder if its complexity is to be kept low. In this context the realistic approach is to dynamically perform the partitioning process at the decoder, where more computational resources are typically available. Since the decoder performs motion estimation to generate the side information, the two neighboring motion-compensated key frames thus generated at the decoder can be used to perform the partitioning at the decoder and help obtain the partition map. At the same time, they can also be used to estimate the dependency model at the decoder, thereby helping estimate the appropriate DVC parameters for each partition. The region map and the estimated DVC parameters are now ready to be communicated to the encoder without causing any extra computational load to the encoder in partitioning the frames.

3.3.1.1 Frame Partitioning

The proposed frame partitioning is performed at the decoder by using the two neighboring motion-compensated key frames P_B and P_F at the decoder where and given by

$$P_F(i,j) = \Gamma_F(X_{2i-1}(i,j)) \tag{3-1}$$

$$P_B(i,j) = \Gamma_B(X_{2i+1}(i,j)) \tag{3-2}$$

Here Γ_F and Γ_B represent the backward and the forward motion compensation operators respectively and (i,j) correspond to the pixel location. The regions are constructed in two phases: an initial segmentation phase to compute SAD and generate appropriate the merging criterion (SAD) leading to an iterative region merging phase.

A. Sum of Absolute Difference (SAD) Computation

In the first phase the two motion-compensated key frames (P_B, P_F) are divided into smaller blocks (block size of 8 by 8 is usually considered small). The residual between the co-located blocks in the two motion compensated frames (P_B , P_F) is computed as follows:

$$Z = \sum_{j=1}^{J}\sum_{i=1}^{I}(P_B(\mathrm{i,j}) - P_F(\mathrm{i,j})) \tag{3-3}$$

where Z is the residual value and (i,j) denotes the location of pixel in the considered block.

B. Segmentation (merging block with similar residual)

In the second phase the blocks are iteratively merged. This operation can be considered as a simple content segmentation that groups into a region the part of the frame having SAD lower than certain threshold. The goal of segmentation is to use the residual and divide its range into partitions that represent similar motion. Each partition contains blocks with similar residual values which are subsequently merged together to compose a region. Partitioning of the residual values (see figure 3-4) is performed based on MSE in such a way that the sum of all variances of the partitions is minimized. This corresponds to the minimum variance in each partition, which can be obtained if the partition elements are close and similar. The threshold values $\tau_1, \tau_2, \ldots \tau_L$ (partitions limits, see figure 3-4) and the $\{\mu_1, \mu_2, \ldots . \mu_L\}$ mean values of each partition are thus obtained by minimizing var_total of (3.4) as shown in (3.5).

$$var\_total = \int_{-\infty}^{\tau_1} (x - \mu_1)^2 f(x) + \int_{\tau_1}^{\tau_2} (x - 2\tau_1 + \mu_1)^2 f(x) + \cdots \int_{\tau_{L-1}}^{\tau_L} (x - \qquad (3\text{-}4)$$

$$2\tau L{-}1{+}2\tau L{-}2{-}2\tau 1{+}\mu 12 f(x)$$

$$var\_total_{opt} = \underset{(\tau_1, \tau_2, \ldots \tau_L)}{\operatorname{argmin}} \; var\_total(\tau_1, \tau_2, \ldots \tau_L) \qquad (3\text{-}5)$$

Solving (3-5) gives the values of $\tau_1, \tau_2, \ldots \tau_L$.

Having obtained the thresholds $\tau_1, \tau_2, \ldots \tau_L$, the pixels are merged together to form partitions. After assigning the side information blocks into L regions $\{X_1, X_2, \ldots X_L\}$, the decoder generates an efficient representation of the partition map and transmits that to the encoder to partition the original WZ frame into corresponding regions.

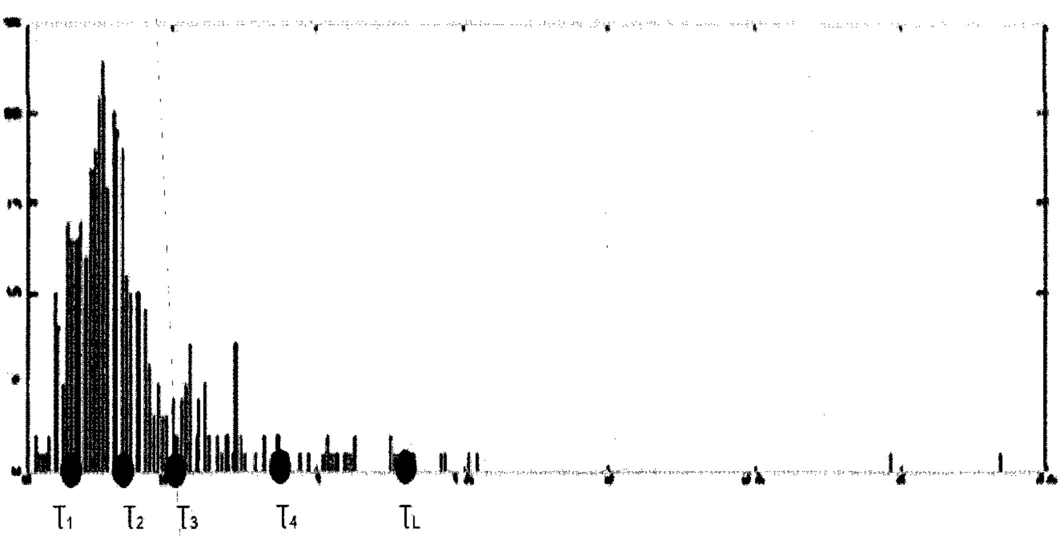

Figure 3-4: SAD histogram and threshold estimation

C. Efficient Region Map Representation by Quadtree

Straight forward coding of the region map to obtain a useful representation is not very efficient. Since permutations of the region numbering do not affect the coding process, we, therefore, consider a transformation of the region map into a quadtree. The use of a quadtree is ubiquitous in image coding [77-79] because of its simplicity and efficiency in coding the block configuration, requiring only 1 bit per node of the tree. QT decomposition is a powerful technique which divides the image into 2-D homogeneous, regions with similar MSE or SAD and thus segments it. The decomposition builds a tree.

Each tree node has four children and it is associated with a uniquely defined region of the image. It is obvious that the root is associated with the whole image. QT decomposition can be done either by top-down or bottom-up procedures. When QT decomposition is used for image compression, the resulting tree is coded. The coding procedure includes coding of the tree structure information and coding of the leaf information. For an illustration Figure 3-5 shows an example of frame partitioning into 4 regions, "1" assigned to the parent node and "0, 2, 4 or 6" to the leaves. Table 3-2 shows the corresponding quadtree structure.

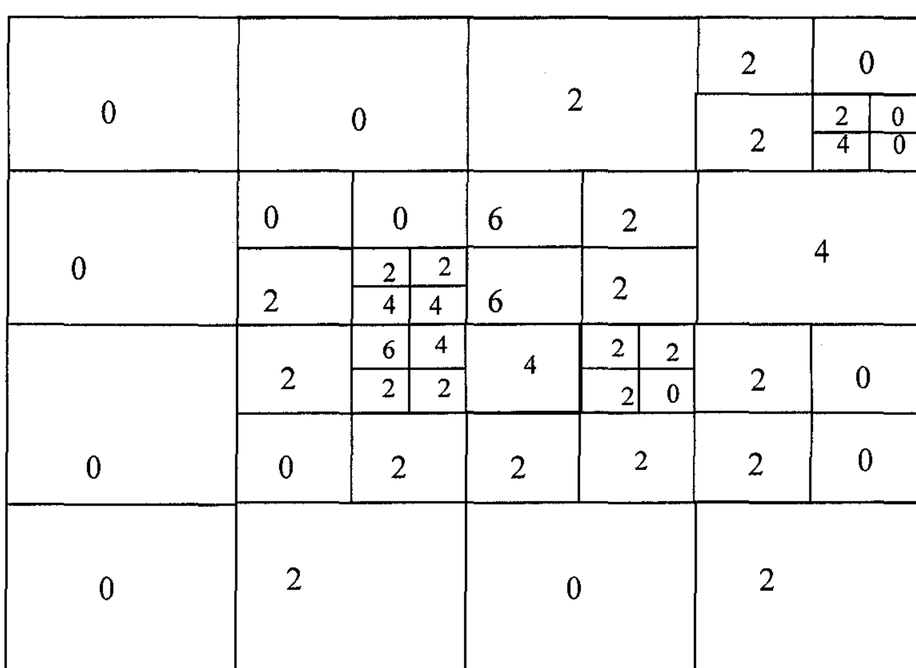

Figure 3-5 Frame Partitioning

Table 3-2 an example of a quad-tree structure

<table>
<tr><td colspan="4">1</td><td colspan="6">1</td><td colspan="4">1</td><td colspan="6">1</td><td>Level 2</td></tr>
<tr><td>0</td><td>0</td><td>0</td><td>1</td><td colspan="2">2</td><td colspan="2">1</td><td>4</td><td>1</td><td colspan="2">1</td><td colspan="2">1</td><td>0</td><td>2</td><td>0</td><td>1</td><td>0</td><td>2</td><td>Level 3</td></tr>
<tr><td>0</td><td>0</td><td>2</td><td>1</td><td>2</td><td>2</td><td>0</td><td>1</td><td colspan="2">6262</td><td>4</td><td>1</td><td>2</td><td>2</td><td colspan="2">2020</td><td>2</td><td>1</td><td>0</td><td>2</td><td>Level 4</td></tr>
<tr><td colspan="4">2424</td><td colspan="6">2040</td><td colspan="4">2220</td><td colspan="6">6422</td><td>Level 5</td></tr>
</table>

3.3.1.2 Dependency Channel Modeling in Pixel Domain

The dependency channel is used to represent the correlation between corresponding pixels in l^{th} region of a given frame $X_l$and its side information Y_l and is described by a Laplacian distribution, as in [57]. The said Laplacian model is characterized by a parameter α_l . This Laplacian parameter is estimated online by using residual statistics between co-located regions in the two motion-compensated key frames $P_{F,l}, P_{B,l}$ regions as follows.

$$Z_l(i,j) = \frac{(P_{Fl}(i,j) - P_{Bl})}{2} \tag{3-6}$$

$$\sigma_l^2 = E_Z \left|\left(Z_l(i,j)\right)^2\right| - \left(E_Z[Z_l(i,j)]\right)^2 \tag{3-7}$$

where $\boldsymbol{\sigma_l^2}$ is a confidence measure of the side information which expresses how good the frame interpolation outcome is and $\boldsymbol{E_Z(Z)}$ is the expectation operation over the residual region.

The Laplacian distribution parameter, $\propto_l$ defined by

$$\propto_l = \sqrt{\frac{\sigma_l^2}{2}} \tag{3-8}$$

is thus calculated by evaluating σ_l^2 online.

3.3.1.3 Quantizer Selection for Region-Based DVC in Pixel Domain

The second parameter to be adaptively selected for each region is depth of the quantizer. As a result, this process generates a matrix of quantizer set for each WZ frame. The quantizer set selection is to reduce the overall distortion to maximize the quality (PSNR) while adhering to rate constraints. The rate-distortion characteristics of each region are computed at the decoder based on a statistical model that is used to estimate the correlation between the original source information and the side information.

To select the best quantizer set for each WZ frame, first an estimate is obtained of the reduction in distortion and the corresponding rate for each region with respect to each quantizer. Then the quantizer set that gives the most distortion reduction ΔD_{WZ} and overall rate less than R_{WZ} is chosen in accordance with:

$$\Delta D_{opt} = \underset{(Q_1, Q_2, \ldots Q_L)}{\operatorname{argmax}} \Delta D_{WZ}(Q_1, Q_2, \ldots . Q_L) \tag{3-9}$$

Subject to $R \geq R_{WZ}$

One can also define the overall bitrate and the proposed system to select the best quantizer set that gives the largest distortion reduction while maintaining the rate constraints for each region. The decoder, after assigning appropriate quantizer to each region transmits the set of quantizers for each region at the WZ frame to the encoder through the feedback channel.

A. Bitrate Estimation

The bitrate for each quantized pixel value is the numbers of bits that are needed to Turbo decode the quantized pixel. Since each bit plane is separately decoded, the bitrate estimation is performed at the bitplane level. Assume Wyner-Ziv region to be encoded is denoted by X_l, and the generated side information is denoted by Y_l. Also assume that the most significant $k-1$ bits of the pixel value $x \in X_l$ have already been error free decoded. Note that the encoder and the decoder know from already decoded $k-1$ bits, $\{b_1, b_2, \ldots . b_k\}$, that x is in the interval $[X_L \;\; X_U]$, where X_L and X_U are lower and upper limits to the x. For example, when $k=1$, $X_L = 0$ and $X_U = 255$. At the encoder, the bit value reduces the interval $[X_c \;\; X_U]$ in such way that $x \in [X_L \;\; X_C]$ if $b_k = 0$ and $x \in [X_C + 1 \;\; X_U]$ if $b_k = 1$ where $X_C = \frac{X_L + X_U}{2}$ An error in b_k occurs, when pixel,x, lies within X_L and X_C i.e., $x \in [X_L \;\; X_C]$ and the corresponding estimated pixel value , y, lies within X_C and X_U, i.e., $x \in [X_C + 1 \;\; X_U]$. Or, similarly, an error occurs when $x \in [X_C + 1 \;\; X_U]$ and the corresponding estimated pixel $y \in [X_L \;\; X_C]$. Assuming that the difference between the original source information and the side information is Laplacian distributed, the conditional pdf of y given x and $X_L \leq y \leq X_U$ is

$$p(y|x, X_L \leq y \leq X_U) = \begin{cases} \dfrac{\frac{\alpha e^{-\alpha|x-y|}}{2}}{p(x, X_L \leq x \leq X_U)} & if \quad X_L \leq x \leq X_U \\ 0 & otherwise \end{cases} \tag{3-10}$$

The parameters α can be estimated as in equation (3-8), the original source information is modeled as random variable $(\acute{X})$, with Laplacian distribution similar to the one for Z.

$$\acute{X} = Y + Z \tag{3-11}$$

From (3-12), the error probability of bit value b_k of pixel x is estimated using

$$P_e(b_k) = \begin{cases} \displaystyle\int_{X_C+0.5}^{X_U} p(y|x, X_L \leq y \leq X_U)\, dy & if \quad b_k = 0 \\ \displaystyle\int_{X_L}^{X_C+0.5} p(y|x, X_L \leq y \leq X_U) dy & if \quad b_k = 1 \end{cases} \tag{3-12}$$

The integration intervals are expanded by 0.5 in order to cover the whole interval$[X_L, X_U]$. Finally the average error probability, P_k for the entire bitplane k is estimated using

$$P_x = \sum_{x=0}^{x=255} p(x) P_e(b_k) \tag{3-13}$$

where the pdf $p(x)$, is obtained from the histogram of the region, which provides the relative frequency of occurrence for each pixel value x. The corresponding rate is obtained by

$$R_k = -P_k \log(P_k) - (1 - P_k)\log(1 - P_k) \tag{3-14}$$

The overall bitrate for the entire region R_l is obtained by

$$R_l = \sum_{k=1}^{K} R_k \tag{3-15}$$

where K is number of bitplanes that represents the bit-depth of the scalar quantizer used to quantize the region pixels.

The overall bitrate for the entire frame R_{WZ} is obtained by

$$R_{WZ} = \sum_{l=1}^{L} R_l \tag{3-16}$$

where index l represents the number of regions.

B. Estimation of Reduction in Distortion

The reduction in the distortion is the difference between the initial distortion and the final distortion after reconstruction. The initial distortion is a distortion between the estimated pixel and the original source pixel. This distortion corresponds to the case when no parity bits are available at the decoder. For the region with (I,J) pixels this distortion is obtained from

$$D_l = \sum_{i=1}^{I}\sum_{j=1}^{J}(X_l(i,j) - Y_l(i,j)) \tag{3-17}$$

To compute the reduction in the distortion, first the distortion between the reconstructed pixel and the original source pixel must be computed, when certain number of parity bits has been obtained from the encoder at some R_l bitrate to correct the estimation error. The value of R_l is estimated as shown in subsection A of section 3.3.1.3. The pixel value of the reconstructed region $\hat{X}_l$ when bitrate R_l is used can be estimated by using the reconstruction method as in (3-18):

$$X = \begin{cases} \left\lfloor \frac{Y}{Q} \right\rfloor \times Q + Q & if \quad \left\lfloor \frac{X}{Q} \right\rfloor > \left\lfloor \frac{Y}{Q} \right\rfloor \\ Y & if \quad \left\lfloor \frac{X}{Q} \right\rfloor = \left\lfloor \frac{Y}{Q} \right\rfloor \\ \left\lfloor \frac{Y}{Q} \right\rfloor \times Q & if \quad \left\lfloor \frac{X}{Q} \right\rfloor = \left\lfloor \frac{Y}{Q} \right\rfloor \end{cases} \tag{3-18}$$

Here Q is the quantization step, and is obtained using $Q = \frac{255}{2^{K-1}}$where K is quantizer bit-depth.

The estimated distortion D_l' can now be computed as

$$\acute{D}_l = \sum^{I}\sum^{J}(\acute{X}_l(i,j) - \hat{X}_l(i,j)) \tag{3-19}$$

The reduction in distortion now is

$$\Delta D_l = D_l - \acute{D}_l \tag{3-20}$$

While the overall reduction is obtained from summing the reduction in distortion over all the regions:

$$\Delta D_{WZ} = \sum_{l=1}^{L} \Delta D_l \tag{3-21}$$

3.3.2 Encoding in Region-Based Adaptive Codec in Pixel Domain

The encoder in the proposed Region-Based Adaptive DVC codec in pixel domain performs the following tasks.

- Decompose the received region map.
- Source coding.
- SW encoding

3.3.2.1 Decomposing the Received Region Map

The received Quadtree representation is interpreted to region map at the encoder. Each digit in received Quadtree corresponds to particular 2-D part of the image the size of this 2-D depends on the level of the digit, for example digit in the first level corresponds to 2-D area represents quarter of the entire image, the second level digit represents eighth of the entire image and so on. The areas with similar digits are combined together as a region; the entire frame is partitioned based on this map into regions.

3.3.2.2 Source Coding in Region-Based Adaptive Codec in Pixel Domain

The first step towards encoding a region X_l in the proposed system architecture, depicted in figure 3-2, is the quantization.

- The pixels of each WZ frame of a video sequence X are quantized using uniform scalar quantizer with 2^K levels; the parameters K corresponds to the number of bits needed to map the pixel into one of the quantizer levels 2^K.

- The K can assume any integer value between 1 and 8 corresponding to a number of quantizer levels ranging from 2 to 256. Varying the K values, different RD performance can be found since each K value has a given rate-distortion point associated. Four rate-distortion points were considered in this work performance evaluation in order to allow comparing the RD performance of the region-based DVC codec and the frame-based pixel domain RD results; those rate-distortion points correspond to K values 1, 2, 3, 4, 5.

3.3.2.3 Slepian-Wolf Encoding

After quantizing the region pixels and forming the K bitplane array associated to the region quantized symbols, each bitplane array is then fed into the Slepian-Wolf encoder, starting from the most significant bitplane array. As shown in Figure 3-6, the Slepian-wolf encoder consists of a Turbo encoder and a buffer:

- The Turbo encoder produces a sequence of redundant bits (parity bits) related to the initial data for each bitplane array; the amount of parity bits produced for each bitplane depends on the Turbo encoder rate, i.e. on the ratio of the Turbo encoder output bits (parity bits) per Turbo encoder input bits.
- The produced parity bits are then stored in the buffer, punctured according to a given puncturing pattern, and transmitted upon decoder request via the feedback channel.

Using a parallel concatenation of two identical constituent Recursive Systematic Convolutional (RSC) encoders, as proposed in [21]. In between the *RSC* s, a variable length-bit random interleaver is employed to de-correlate the variable-length input sequence *(X)* between the two RSC encoders [21]. In Figure 3-6, the $\mathbf{S}_i$ and $\mathbf{P}_i$ (i = 1, 2) symbols represent the sequences produced by the RSC_i encoder, C_1 and C_2 respectively.

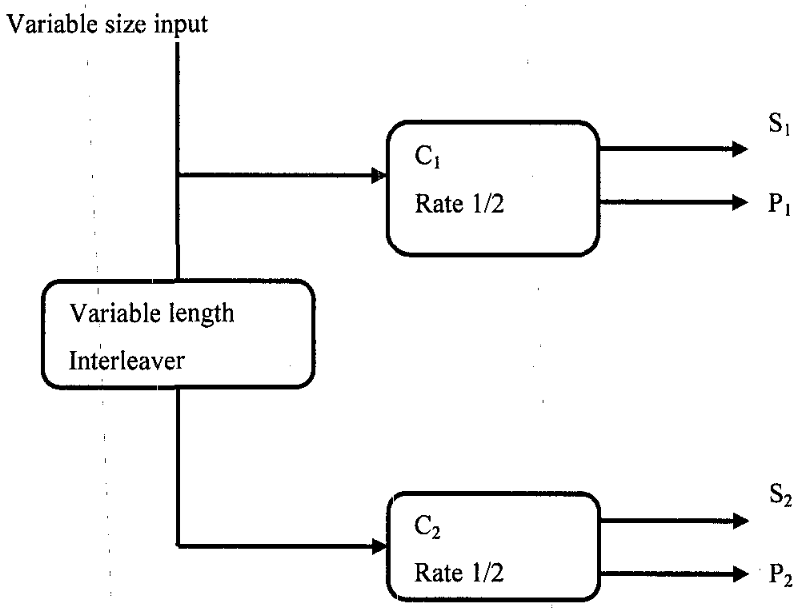

Figure 3-6: Turbo coder structure

A. Variable length random interleaver:

One of the Turbo encoder architectural modules is the interleaver

- The interleaver function is to rearrange its input sequence and its input sequence in a different order according to a given pattern, the interleaving structure; the order of the interleaver input sequence is restored by feeding the interleaver output sequence into a de-interleaver.
- The interleaving structure maps each index of the input sequence into one index of the output sequence, in a univocal correspondence.
- In the Region-based Adaptive DVC codec, a pseudorandom interleaver is adopted in order to allow comparing the Region-based performance with the pixel domain frame based solution in [11]. Note that the input sequence cannot be interleaved in a completely random fashion in order to make possible the Turbo decoding. Since, the Turbo decoder needs to know the interleaving pattern used by the Turbo encoder to be able to perform the decoding task.
- The interleaver length also affects the Turbo coding performance [80]. As shown by Shannon, random codes with large block sizes may achieve a transmission rate close to the channel capacity (amount of bits that can be reliably transmitted, i.e. with a bit error probability near zero, over the channel) [80].

B. RSC Encoder

Beside the interleaver as shown in Figure 3-6, the Turbo encoder includes two recursive systematic convolutional (RSC) encoders. A RSC encoder is typically characterized by a generator matrix G which allows obtaining the RSC encoder output C_{out} for a given RSC encoder input. The RSC encoder output C_{out} is generated according to (3-22).

$$C_{out} = C_{in} \times G \tag{3-22}$$

In the [11, 65] solutions, RSC encoders of rate ½ were used. These RSC encoders of rate ½ employed in [13] are used as benchmark in this thesis. The constituent recursive systematic convolutional encoder of rate ½ is represented by a generator matrix of the form $\begin{bmatrix}1 & \frac{g_2(Đ)}{g_1(Đ)}\end{bmatrix}$, where Đ is the delay symbol. Figure 3-7 illustrates a rate ½ constituent recursive systematic convolutional encoder with memory $m = 4$ (16 states) and with a generator matrix given by (3-23)

$$\begin{bmatrix}1 & \frac{1 + Đ + Đ^3 + Đ^4}{1 + Đ^3 + Đ^4}\end{bmatrix} \tag{3-23}$$

From Figure 3-7 we can see that the rate of the RSC encoder is ½ ; for each input bit there are two output bits, these two output bits at time n depends on the input at the same time and the values in shift registers (Đ) at time $n - 1$. The set of values $(s_1, s_{2,}, s_3, s_4)_{n-1}$ represents the RSC encoder state S at time $(n - 1)$, i.e. $S_{n-1} = (s_1, s_{2,}, s_3, s_4)_{n-1}$. After introducing the n^{th} input bit, the shift registers a value change accordingly S_{n-1} to S_n (the RSC code state at time n). Figure 3-7 shows that, for each RSC state, the bits u_n shifted into the RSC encoder and the corresponding RSC output bits, u_n^s and u_n^p (see Figure 3-7).

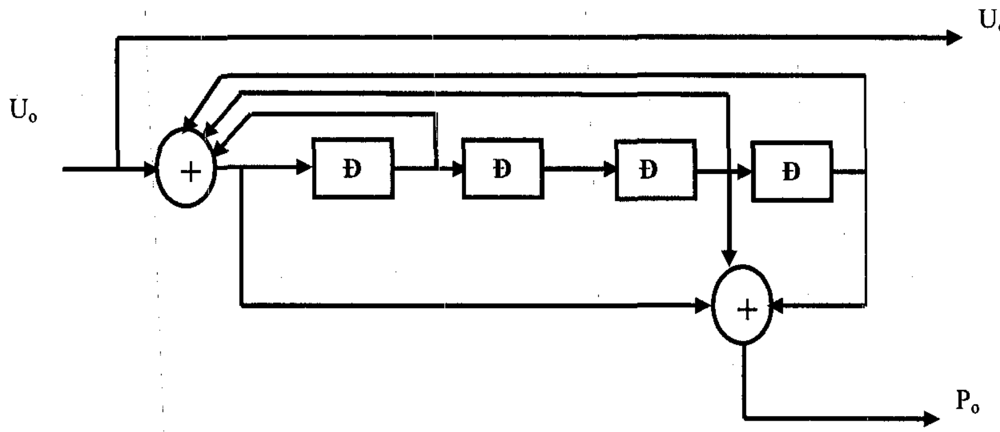

Figure 3-7: RSC encoder structure

Each one of the two RSC encoders enclosed by the Turbo encoder computes a parity sequence corresponding to the O-bit sequence at its input. For the C_1 encoder, the input sequence is a bitplane extracted from the quantized symbol stream (see Section 3.3.2.1); for the C_2 encoder, the input sequence is a pseudo-randomly interleaved version of the bitplane (see Figure 3-7). The input sequence of each RSC encoder is reproduced since the convolutional encoders are systematic. The reproduced input in the output sequence is known as the systematic sequence and is represented in Figure 3-6 by S_1 and S_2 (for the C_1 and the C_2 encoders, respectively). The Turbo encoding output consists of the systematic sequences S_1 and S_2 and the parity sequences, P_1 and P_2. The systematic sequences S_1 and $S_2$are discarded, as in [11] (in practice, this is the information estimated with the side information at the decoder); on the other hand, the parity sequences, P_1 and P_2, are stored in the buffer.

C. Rate Compatible Punctured Turbo (RCPT) Code-Based Slepian-Wolf Codec

The Slepian-Wolf codec rate control technique is built based on a Rate Compatible Punctured Turbo (RCPT) code structure [81]. The process of rate-compatibility restriction organizes the rate hierarchically, where higher rate codes are embedded in the lower rate codes. The resulting code is known as the Rate compatible punctured Turbo Code (RCPTC) scheme, as in [11]. The RCPT is implemented by dividing the *O-bits* parity sequence generated by each RSC encoder into packets of P bits (O/ρ), where O is the interleaver length and ρ is the puncturing period. The decoder requests parity bits via the feedback channel, the Wyner-Ziv encoder only transmits a packet of (O/ρ) bits to the decoder; of course, no packet is sent twice. When the decoder requests for more parity bits, only some of the O bits parity sequence is sent to the decoder, corresponding to a (O/ρ)-bit packet, puncturing is performed only over the parity sequence because the systematic sequences are discarded . When the similarity between the region to code and its estimation made at the decoder is high, a few of the parity bits (Wyner-Ziv bits) are needed to be sent to the decoder to reach a certain quality (or to successfully decode the quantized symbol stream). The puncturing pattern, i.e. the order by which the blocks of (O/ρ) bits (both of P_1 and P_2) are transmitted to the decoder, is based on the pseudo-random puncturing pattern, the blocks are transmitted alternately from P_1 and P_2. The first packet of the parity bits is chosen from sequence P_1; no parity packet is sent from the C_2 parity sequence (P_2). The next packet if a request is made via the feedback channel is sent from P_2. The decoder requests for more parity bits packet via the feedback channel, the encoder sends packets from P_1 for the odd number of request and from P_2 if the number of request, is even and so on until no more requests are made, or parity bit packets are sent.

3.3.3 Decoding in Region-Based Adaptive DVC Codec in Pixel Domain

The decoder in the proposed Region-Based Adaptive DVC codec performs the following tasks

- Side information generation.
- SW decoding.
- Reconstruction.

3.3.3.1 Side Information Generation

The motion estimation techniques, attempt to choose the best prediction for the current frame in the rate-distortion sense used in traditional video coding at the encoder. In other words, for a given block in the current frame, the motion estimation techniques attempt to find the best match in the reference frame independently of the true motion of the block in the scene. For frame interpolation, the current frame is not known and it is necessary to estimate the true motion to correctly interpolate the missing frame (current frame), these techniques are not well-suited since in this case, usually by motion compensation between temporally adjacent frames. In the existing DVCs [11, 37, 38, 40, 51], motion-compensated interpolation is adopted to obtain more accurate side information as compared to other methods, such as previous frame duplication and non-motion compensated interpolation. The method used in this work is based on side estimation B-frame (SE-B) as proposed in [11].

The main idea of this approach is to model the motion vectors for the current WZ frame using the motion vectors calculated in the previous frames. By using 3D recursive block matching attempts are made to find the true motion field. The block-based motion estimation algorithm was chosen due to its low complexity comparing to other algorithms such as dense motion vector fields or motion segmentation with arbitrary regions of support. As depicted in figure 2-7, the steps of creating the side information is as follow

1. The forward motion vectors MV_F estimation: - A block matching algorithm is employed to estimate the motion between the next X_{2i-1} and previous X_{2i+1} temporally adjacent key frames.

$$MV_F = \vee\ (X_{2i-1}, X_{2i+1}, w, r, \Delta) \qquad (3\text{-}24)$$

$\vee \equiv$ Operator represents the 3-D motion estimation process.

$w \equiv$ Dimension of the square block where the motion search is performed within.

$r \equiv$ The search radius parameter, it defines the X_{2i-1} area dimension in which the block most similar to the current block in the X_{2i+1} frame is searched for.

$\Delta \equiv$ The step size in pixels. It is a value by which a fixed window size used in searching for MV in previous frame X_{2i-1} is stepped by. This parameters helps in reducing the computational complexity.

2. The backward MV_B estimation: - The same as the forward motion vectors estimation procedure, except the reference frame is the previous frame X_{2i+1} and next frame as the source X_{2i-1}

$$MV_B = \vee\ (X_{2i-1}, X_{2i+1}, w, r, \Delta) \qquad (3\text{-}25)$$

3. The first estimation is obtained by applying $\frac{MV_F}{2}$ on the previous frame X_{2i+1} as in

$$P_F = \Gamma(\frac{MV_F}{2}, X_{2i+1}) \tag{3-26}$$

where Γ the motion compensation operator as is defined in 3.3.1.1

4. The second estimation is obtained by applying $\frac{-MV_B}{2}$ on the next frame X_{2i-1}

$$P_B = \Gamma(\frac{MV_B}{2}, X_{2i-1}) \tag{3-27}$$

5. The final side estimation frame is obtained as the mean between P_F and P_B.

$$Y = \frac{P_B + P_F}{2} \tag{3-28}$$

In fact, it models the motion vectors for the current frame X as half the motion vector for the next frame P_F with P_B as the reference.

3.3.3.2 Slepian-Wolf Decoder

The main goal of the Slepian-Wolf decoder module is to estimate each bitplane extracted from the quantized symbol stream. The Slepian-Wolf decoder uses the received part of the parity sequences, P_1 and P_2, sent by the WZ encoder (see Figure 3-6) and the available side information or a "noisy" version of the encoded frame X_{2i} generated at the decoder. The Slepian-Wolf decoding is represented by an iterative Turbo decoding procedure, as in [11]. Figure 3-8 depicts a Turbo decoder constituted by two identical Soft-Input, Soft-Output (SISO) decoders; one per each constituent RSC code used in the Turbo encoder implementation. The decoding procedure is performed iteratively since the two SISO exchange the information. The decoding decision is made based on soft information which is the probability, the SISO decoder input and output is soft information. The $SISO_i$ decoder's inputs (*soft* inputs) are represented by the symbols Y, P_i and $L_i^{apriori}$ (i = 1, 2); Y is often known as systematic information, P_i as channel information and $L_i^{apriori}$ as *a priori* information. Since the systematic part of each RSC encoder is discarded, the Wyner-Ziv decoder uses its estimated side information Y, generated through frame interpolation techniques (see 3.3.2), the estimate is seen as a noisy version of X. The noise in the Wyner-Ziv context is the difference between the original X; region and the side information Y for that region. The difference depends on the quality of the estimation at the decoder and it is considered as a channel model.

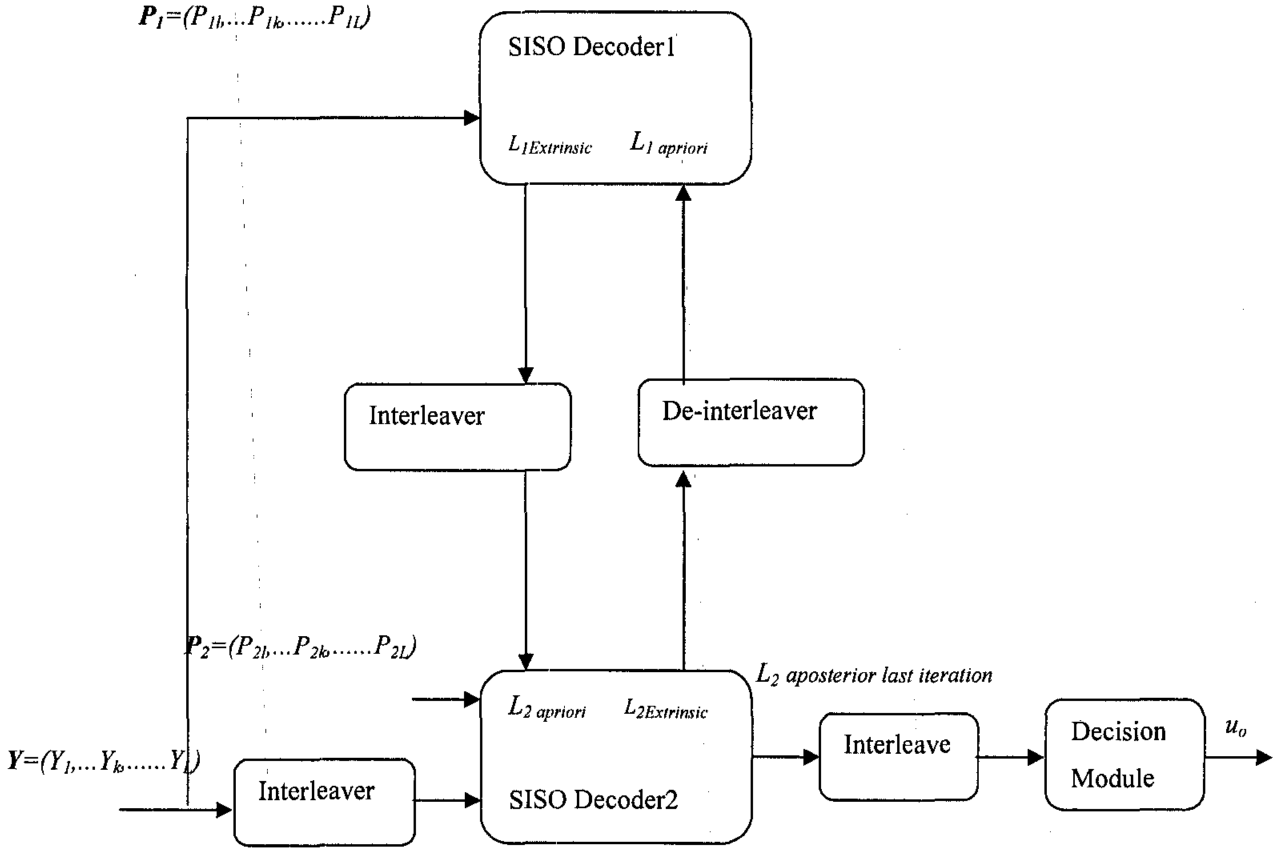

Figure 3-8: Turbo decoder structure

A. SISO Decoding Algorithm

The decoding algorithm used in the SISO decoder implementation is a modified version of the maximum *a posteriori* (MAP) decoding algorithm proposed in 1974 [29]. The iterative SISO decoding procedure is carried out between two modified MAP decoders operating over logarithms of likelihood ratios associated to the systematic and parity bits instead of the bits themselves. The estimate of each bit u_k, denoted by $\acute{u}_n$ is obtained from the Logarithm of the *A Posteriori* Probability (LAPP) ratio defined by (equation (3-29)) [30]. In the symbol-by-symbol MAP decoder, the decoder decides u_n if $P(u_n = +1) > P(u_n = -1)$ and it decides $u_n = -1$ if otherwise. The decision $\acute{u}_n$ is given by

$$L(\acute{u}_n) = ln\left(\frac{P(u_n = +1|r)}{P(u_n = -1|r)}\right) \tag{3-29}$$

Where $P(u_n = \pm 1|r)$ is the a posteriori probability is $L(\acute{u}_n)$ stands for the LAPP ratio and the operator $ln(.)$ is the natural logarithm. The symbol $r_n = (r_1, \ldots . r_O)$ with $r_n = (r_n^S, r_n^P) = (Y_n, P_{in})$ represents the information at the Turbo decoder for each u_n: that is, the parity bit $P_{i,n}$ (i represents the index of the RSC encoder from which P_n belongs to) and the systematic bit Y_n. Incorporating the trellis of the code this may be written as

$$L(\acute{u}_{n)} = ln \frac{\sum_{S^+} P(s_{n-1} = \acute{s},\ s_n\acute{} = s, r)p(r)}{\sum_{S^-} P(s_{n-1} = \acute{s},\ s_n\acute{} = s, r)p(r)} \tag{3-30}$$

Where $s_n \in S$ is the state of the encoder at time n, S^+ is the set of ordered pairs $(\acute{s}, s)$ corresponding to all state transition $s_{n-1} = s \to \acute{s}_n \to s$ caused by the data input $u_n = +1$, and S^- is similarly defined for $u_n = -1$. By canceling the $p(r)$ we only need an algorithm to compute $p(s,'s,r)$, the BCJR for doing this is

$$p(s,'s,r) = \breve{\alpha}_{n-1}(\acute{s})\gamma_n(s;'s)\breve{\beta}_n(s) \tag{3-31}$$

The probability $\breve{\alpha}_n(s_n)$ can be computed from

$$\breve{\alpha}_n(s_n) = \frac{\sum \breve{\alpha}_{n-1}(s)\ \gamma_n(s;'s)}{\sum_{S^+}\sum_{S^-}\sum \breve{\alpha}_{n-1}(s)\,\gamma_n(s;'s)} \tag{3-32}$$

With initial conditions of $\breve{\alpha}_0(0) = 1$ and $\breve{\alpha}_0(s \neq 0) = 0$ this condition states that the encoder is expected to start in stat 0.

The probability $\breve{\beta}_{n-1}(s_{n-1})$ can be computed from

$$\breve{\beta}_{n-1}(s_{n-1}) = \frac{\sum \breve{\beta}_n(s_n)\gamma_n(s_{n-1}, s_n)}{\sum_{S^+}\sum_{S^-}\breve{\beta}_n(s_n)\gamma_n(s_{n-1}, s_n)} \tag{3-33}$$

With boundary conditions $\breve{\beta}_O(0) = 1$ and $\breve{\beta}_O(s \neq 0) = 0$ the encoder is expected to end in state 0 after O input bits, implying that the last m inputs bits, called termination bits are also selected.

From [30], the probability $\gamma_n(s_{n-1}, s_n)$ can be written as

$$\gamma_n(s_{n-1}, s_n) = P(u_n)P(r_n|u_n) \tag{3-34}$$

And the a *priori* probability $P(u_n)$ can be expressed as

$$P(u_n = \pm 1) = \frac{e^{-\frac{1}{2}L^{apriori(u_n)}}}{1 + e^{-L^{apriori(u_n)}}} e^{\frac{u_n L^{apriori(u_n)}}{2}} \tag{3-35}$$

Where $L^{apriori}(u_n)$ is defined by the equation $L^{apriori}(u_n) \approx \frac{P(u_n=-1)}{P(u_n=+1)}$

$$L^{apriori}(u_n) \approx \frac{P(u_n = -1)}{P(u_n = +1)} \tag{3-36}$$

In equation (3-34) the $P(r_n|u_n)$ factor is the conditional pdf of (r_n) given that u_n was transmitted; $r_n = (r_n^S, r_n^P) = (Y_n, P_{in})$ with P_{in} representing n^{th} parity bit received at the decoder (i represents the index of RSC encoder from which P_n belongs to) and the Y_n the n^{th} systematic bit for each u_n.

B. Modeling the Residual of the Parity Information

In order to compute the probability $\gamma_n(s_{n-1}, s_n)$, it is necessary to know how to model the conditional pdf $p(r_n|u_n)$. In [65], it is assumed that no errors are introduced in the parity bits (i.e. the Wyner-Ziv bitstream) transmission. As it is well-known, the error-free transmission scenario is the most favorable transmission scenario that can be taken into account and therefore it is in this situation that the best Wyner-Ziv codec RD performance can be achieved. Since no errors are introduced during the parity bitstream transmission, the parity bits transmitted are equal to the parity bits received at the decoder (generically represented in Figure 3-8 by P_1 and P_2).

As was mentioned in Section 3.1.2, puncturing is used at the encoder and therefore some parity bits are not transmitted to the decoder; those bits are often called deleted bits. Since the decoder knows the puncturing pattern (Section 3.6.2.3), the deleted bit positions in the bitstreams P_1 and P_2 are known. The deleted bit positions are filled with a zero value, indicating that no value was received for that parity bit position [30]; remember that, in the P_1 and P_2 bitstreams, the value $+1$ corresponds to the bit 1 and the value -1 corresponds to the bit 0. In this work the parity is assumed to be transmitted over a Gaussian channel where the Gaussian distribution variance σ^2 is arbitrarily small.

C. Modeling the Conditional pdf $p(r_n|u_n)$.

The probability $p(r_n|u_n)$ can be written as

$$p(r_n|u_n) = p(r_n^S|u_n) \times p(r_n^P|u_n) \tag{3-37}$$

$$= e^{-\alpha|r_n^S - u_n^S|} \times e^{\frac{-\left(r_n^P - u_n^P\right)^2}{2\sigma^2}} \tag{3-38}$$

$$= e^{-\alpha|r_n^S - u_n^S|} \times e^{\frac{-\left(r_n^P\right)^2 + \left(u_n^P\right)^2 - 2\times r_n^P \times u_n^P}{2\sigma^2}} \tag{3-39}$$

$$= B_n \times e^{-\alpha|r_n^S - u_n^S|} \times e^{\frac{r_n^P \times u_n^P}{\sigma^2}} \tag{3-40}$$

The "error" or residual between the Turbo encoder input sequence (Turbo encoder systematic information) and the corresponding information at the Turbo decoder (Y) is modeled by Laplacian distribution, the parity information error is modeled by a Gaussian distribution with an arbitrarily small variance.

$$\gamma_n(s_{n-1}, s_n) = A_n \times e^{-\frac{u_n}{2}L^{apriori}} \times B_n \times e^{-\alpha|r_n^S - u_n^S|} \times e^{\frac{r_n^P \times u_n^P}{\sigma^2}}$$

$$\gamma_n(s_{n-1}, s_n) = e^{-\frac{u_n}{2}L^{apriori}} \times e^{-\alpha|r_n^S - u_n^S|} \times e^{\frac{r_n^P \times u_n^P}{\sigma^2}} \tag{3-41}$$

Since the term $\gamma_n(s_{n-1}, s_n)$ appears in both the numerator and the denominator of (3-42) the product $A_0 \times B_0$ is independent of u_o and therefore can be cancelled.

Logarithm of the *a Posteriori* Probability (LAPP)

$$L(\acute{u}_n) = ln\left[\frac{\sum_{S^+} \breve{\alpha}_{n-1}(s_{n-1}) e^{-\frac{u_n}{2}L^{apriori}} \times e^{-\alpha|r_n^S - u_n^S|} \times e^{\frac{r_n^P \times u_n^P}{\sigma^2}} \times \breve{\beta}_n(s_n)}{\sum_{S^-} \breve{\alpha}_{n-1}(s_{n-1}) e^{-\frac{u_n}{2}L^{apriori}} \times e^{-\alpha|r_n^S - u_n^S|} \times e^{\frac{r_n^P \times u_n^P}{\sigma^2}} \times \breve{\beta}_n(s_n)}\right] \tag{3-42}$$

$$L(\acute{u}_n) = L^{apriori}(u_n) + ln\left[\frac{\sum_{S^+} \breve{\alpha}_{n-1}(s_{n-1}) e^{-\alpha|r_n^S - u_n^S|} \times e^{\frac{r_n^P \times u_n^P}{\sigma^2}} \times \breve{\beta}_n(s_n)}{\sum_{S^-} \breve{\alpha}_{n-1}(s_{n-1}) \times e^{-\alpha|r_n^S - u_n^S|} \times e^{\frac{r_n^P \times u_n^P}{\sigma^2}} \times \breve{\beta}_n(s_n)}\right] \tag{3-43}$$

D. Converting the Side Information into Soft Information

The SISO's input (Y) is computed as a ratio between the case of $u_n = +1$ and $u_n = -1$; which is corresponding to $[X_L \quad X_U]$ the upper and lower limits to the pixel value for the first bitplane [0 255] at the pixel domain implementation, for the transform domain the upper and lower limits are first found at the encoder and transmitted to the decoder. Applying the successive re-finable theorem (section 2.1.6) to find the corresponding $[X_L \ X_U]$ for the next bitplane, the bit value u_k shrinks this interval in such a way that $X \in [X_L \ X_C]$ if $u_n = +1$ and $X \in [X_C + 1 \ X_U]$ if $u_n = +1$ where o represents the index of the bitplane $n = \{1, 2, \ldots, O\}$.

E. Decision Operation

After the iteration process stops, the last iteration *a posteriori* information of $SISO_2$ is computed using the equation (3-42). The estimate $\acute{u}_o$ of each Turbo encoder input bit u_n results from thresholding the $SISO_2$. Since the transmitted symbols are $+1$ and -1, the threshold value is set to zero (central value between +1 and −1). After having the estimate $\acute{u}_n$ for each u_n, the bit error rate P_e, is calculated. If P_e is greater than a given bit error rate threshold, the Slepian-Wolf decoder requests more parity bits from the Slepian-Wolf encoder via the feedback channel and the iterative decoding procedure continues; otherwise, the Turbo decoding procedure ends and the output of the Turbo decoder is the sequence of $\acute{u}_n$s from $n = 1$ to $n = O$ which corresponds to an estimate of a given bitplane).

F. Estimation of the Residual Error Probabilities { P_e }

The comparison between each estimate $\acute{u}_n$ and each information bit u_n which is often called ideal error detection is not possible, since in practice the original bitplane is not available at the decoder. Alternatively the P_e is computed as [82]

$$P_e = \frac{1}{O} \sum_{n=1}^{O} \frac{1}{1 + e^{|L_n|}} \tag{3-44}$$

where O is the length of the bitplane, L_n is the log likelihood ratio from (3-44) for n^{th} bit in the considered bitplane.

3.3.3.3 Reconstruction in the Region- Based Pixel Domain Codec

The last stage in the decoding process is the Wyner-Ziv region reconstruction.

- After decoding K bitplanes associated to the last region quantized pixel values, these K bitplanes are grouped to form the decoded quantized symbol stream. Notice that the decoder knows the $o$parameter value used at the encoder side since it is transmitted to the decoder in a bitstream header.
- The decoder rearranges the regions back based on the quadtree.
- Once the decoded quantized symbol stream $\hat{q}_l$ is obtained, the reconstruction of the X_l region can be performed.
- The region-based codec performs the reconstruction process as in [11, 65],which can be described by one of three cases as following:

  I. Case 1: when the side information value is smaller than the lower bound of the decoded quantization interval, the lower bound of the quantization interval of the decoded quantization index is taken as the reconstructed pixel value.

  II. Case 2: when the side information is inside of the quantization interval, the current side information value is used as a reconstructed pixel value.

  III. Case 3: when the side information value is greater than the upper bound of the decoded quantization interval, the upper bound of the decoded quantization interval is taken as the reconstructed pixel value.

3.4 Region-Based Adaptive DVC codec in Transform Domain

The main differences between the Region-Based Adaptive DVC codec in the transform domain and Region-Based Adaptive DVC codec in Pixel domain solutions are in the DCT/IDCT, quantization and reconstruction modules as shown in figure 3-3 This system in the light of the higher level abstraction architecture Figure 3-1 contains the following parts

- The Frame partitioning and DVC parameters.
- The Encoder: - Source coding and SW encoding.

– The Decoder: - Side information generation, SW decoding and reconstruction.

3.4.1 Frame Partitioning and DVC Parameters

The partitioning is performed in pixel domain hence, the technique of frame partitioning is similar to the proposed Region-Based Adaptive DVC codec pixel domain solution 3.3.1.1.

3.4.1.1 Channel Modeling in Transform Domain

After the decoder partitions the side information into regions, the two motion-compensated key frames are also partitioned into regions based on the constructed regions map. Each co-located regions in the two motion compensated key frames are first transformed into a DCT domain. The residual statistics between the corresponding regions coefficients in $\boldsymbol{X}_l^{DCT}$ and $\boldsymbol{Y}_l^{DCT}$ is assumed to be modeled by a Laplacian distribution, as in [13, 65]. The Laplacian parameter is estimated online for the region at the DCT band level, i.e. each DCT band has a Laplacian parameter associated by using the corresponding regions in the two motion-compensated key frames $\boldsymbol{P}_B^l$ and $\boldsymbol{P}_F^l$ by using equation (3-7) and equation (3-8) $\boldsymbol{Z}_l^{DCT}$ is obtained by

$$Z_l^{DCT}(i,j) = DCT\big(Z_l(i,j)\big) \tag{3-45}$$

$$\sigma_l^2 = E_Z\big|[Z_l^{DCT}(i,j)]^2\big| - (E_Z[Z_l^{DCT}(i,j)])^2 \tag{3-46}$$

$$\alpha_l^{DCT} = \sqrt{\frac{\sigma_l^2}{2}} \tag{3-47}$$

3.4.1.2 Quantization Matrices Selection for Region in Transform Domain

The rate-distortion characteristics of each region are computed at the decoder based on a statistical model that is used to estimate the correlation between the original source DCT bands and the side information DCT bands. Each region is quantized adaptively, i.e, for each region of the WZ frame one quantizer matrix is chosen from 8 quantizer matrices of Figure 3-10, based on its rate-distortion characteristics. The quantizer matrices set selection is to reduce the overall distortion to maximize the quality (PSNR) while adhering to rate constraints. To select the best quantizer matrices set for each WZ frame, first the distortion reduction and the corresponding rate for each region with respect to each quantizer is computed. The process of selecting the appropriate quantizer matrices set in transform domain is similar to quantizer set selection in pixel domain in section 3.3.1.3

3.4.2 Encoding in Region-Based adaptive DVC codec Transform Domain

The encoder in the proposed Region-Based adaptive DVC codec in transforms domain performs the following tasks

–DCT

–Source coding

–Slepian-Wolf encoding

A block-based DCT is employed in the Region-Based Adaptive DVC architecture with the same purpose as in traditional video coding schemes: to de-correlate block samples by exploiting the spatial redundancy between neighboring samples and to compact the block energy into as few transform coefficients as possible.

3.4.2.1 Source Coding In the Transform Domain

After the DCT of a given region is performed at the encoder, these DCT coefficients are grouped in bands, all DCT coefficients located at the same index into one band B_k. The DC coefficients and AC coefficients of all the blocks of the region are read row wise, while the blocks themselves are read zigzag and quantized. The quantized coefficient values are then subsequently read per bit-plane.

A. DC Coefficient Quantization

The DC coefficient band is characterized by high amplitude positive values since each DC transform coefficient expresses the average energy of the corresponding 4×4 samples block. Since only positive values are fed into the quantizer, the quantization algorithm for the DC coefficient band can be similar to the one used in the pixel domain implementation.

B. AC Coefficients Quantization

For the remaining DCT coefficient bands, called AC coefficients bands, the quantizer input assumes both positive and negative values since the basic functions associated to the AC coefficients present zero mean values [83].

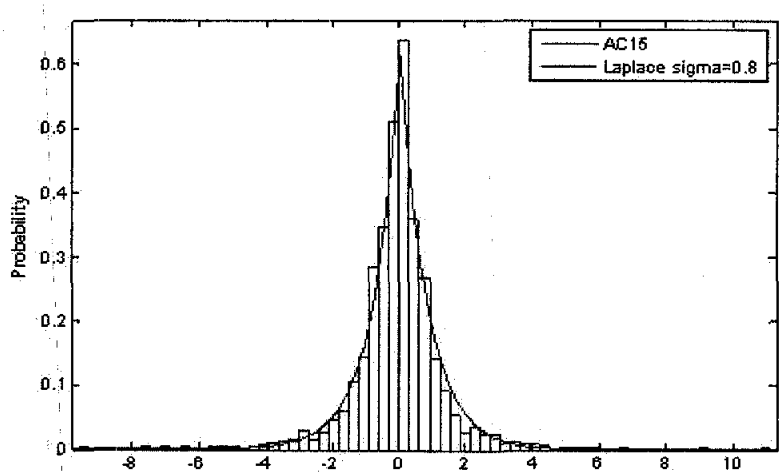

Figure 3-9 DCT coefficient (AC 15) foreman video sequence

Figure 3-9 depicts the AC coefficients distribution for the band B_t, corresponding to the lowest and the highest spatial frequency AC bands, respectively, of the *Foreman* QCIF video sequence. As it can be observed, the AC coefficients distribution is rather symmetrical around the zero amplitude, this DCT coefficients distribution characteristic occurs not only for the band B_t but also for all the AC bands. Considering the distribution of AC coefficients bands, it is well-suited to use a quantizer with a quantization interval symmetric around zero; low DCT coefficient values around zero are now quantized under the same quantization interval index (independently of its signal) avoiding errors between them X_l^{DCT} and Y_l^{DCT} corresponding quantized symbols and therefore reducing the annoying block artifact effect.

The amplitude of the AC coefficients, within a 4×4 DCT coefficients block, tends to be higher for the coefficients close to the DC coefficient and to decrease as the coefficients approach the higher spatial frequencies. In terms of AC coefficient bands, this means that AC bands labeled with indexes closer to 1 (index of DC band) enclose values of higher amplitude comparing to AC bands labeled with indexes far away from the DC band. In fact, within a 4×4 DCT coefficients block, the lower spatial frequencies enclose more relevant information about the block than the high frequencies, which often correspond to noise or less important details for the human visual system (HVS) [83].

To quantize the coefficients of a certain DCT band B_t , *it* is necessary to have knowledge of its dynamic range, i.e. in which range the DCT coefficients vary, besides the number of quantization levels associated to that band; the quantization step size, necessary to define the quantization intervals bounds, can be computed from these two quantities. The dynamic range for each DCT band is transmitted frame by frame to the decoder and assumed to be error-free when received to allow having quantization interval widths adjusted to the dynamic range of each band.

The quantization step size for the DCT band B_t results from the dividing the range of the B_t by number of quantization levels 2^{K_t} . The range for the DC band is $[0,\ 2^{11}]$ [64], hence, the range of the quantization interval is given by

$$I_{DC}^{q}=[q \times W_{DC}, (q+1) \times W_{DC}] \tag{3-48}$$

$$q = \frac{V_{DC}}{W_{DC}} \tag{3-49}$$

where W_{DC} is the DC quantization step size and given by

$$W_{DC} = 2^{11-K_t} \tag{3-50}$$

To obtain the quantization step size for the AC bands B_t, $t = 2, \ldots, 16$, the highest absolute value within each band B_t is first determined $|V_t|_{max}$. Since this computation is performed over absolute values, the highest absolute value determined corresponds to half of the total B_t band range, therefore the range for the AC band B_t is$[-|V_t|_{max}, +|V_t|_{max}]$. The quantization step size W is then obtained from (3-51)

$$W_{AC_t} = \frac{2\times|V_t|_{max}}{2^K-1} \tag{3-51}$$

The quantization intervals index q, also known as quantized symbol is given by

$$q = \frac{V_t}{W_{AC_t}} \tag{3-52}$$

where V_t is the DCT coefficient value within the DCT band B_t and W_{AC} is the quantization step size associated to the B_t band.

As in [13], different rate-distortion points are considered to evaluate the codec performance making use of the 4×4 Quantization matrices depicted in Figure 3-10: Quantizer sets for different rate points. The bottom right-most value corresponds to the highest AC coefficient.

$$\begin{bmatrix}16 & 8 & 0 & 0\\ 8 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix} \quad \begin{bmatrix}32 & 8 & 0 & 0\\ 8 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix} \quad \begin{bmatrix}32 & 16 & 8 & 0\\ 16 & 8 & 0 & 0\\ 8 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix} \quad \begin{bmatrix}32 & 16 & 8 & 0\\ 16 & 8 & 0 & 0\\ 8 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$$

$$\begin{bmatrix}32 & 16 & 8 & 4\\ 16 & 8 & 4 & 4\\ 8 & 4 & 4 & 0\\ 4 & 4 & 0 & 0\end{bmatrix} \quad \begin{bmatrix}64 & 16 & 8 & 8\\ 16 & 8 & 8 & 4\\ 8 & 8 & 4 & 4\\ 8 & 4 & 4 & 0\end{bmatrix} \quad \begin{bmatrix}64 & 32 & 16 & 8\\ 32 & 16 & 8 & 4\\ 16 & 8 & 4 & 4\\ 8 & 4 & 4 & 0\end{bmatrix} \quad \begin{bmatrix}128 & 64 & 32 & 16\\ 64 & 32 & 16 & 8\\ 32 & 16 & 8 & 4\\ 16 & 8 & 4 & 0\end{bmatrix}$$

Figure 3-10: Quantizer sets for different rate points

Within a 4×4 quantization matrix, the value at position t, indicates the number of quantization levels associated with the DCT coefficients band B_t. In Figure 3-10, the value 0 means that no parity bits are transmitted to the decoder for the corresponding bands and that the decoder will replace the DCT bands to which no parity bits are sent by the corresponding DCT coefficients bands of the side information determined at the decoder.

After quantizing the DCT coefficients band B_t the quantized symbols (represented by integer values) are converted into a binary stream. The quantized symbols bits of the same significance (e.g. the most significant bit) are grouped together forming the corresponding bitplane array which are then independently Slepian-Wolf encoded.

3.4.3 Decoding in the Region-Based Adaptive Codec Transform Domain

The decoder in the proposed Region-Based Adaptive DVC codec performs the following tasks

- The Side information generation.
- SW decoding.
- Reconstruction.

The process of the side information generation is similar to the proposed Region-Based Adaptive codec in pixel domain side information generation process explained in section 3.3.3.1. The SW decoding for the DCT quantized symbols is done similar to the SW decoding for the pixel quantized symbol explained in section 3.3.3.2, therefore this section will only explain in detail the reconstruction process in transform domain.

3.4.3.1 Reconstruction in the Region-Based Adaptive Codec in Transform Domain

The reconstruction process in the transform domain is performed one DCT coefficient band at a time. The quantization step size for each DCT coefficient band is computed at the decoder by using the highest absolute value of the AC band and the number of the quantization levels that are determined at the encoder and sent to the decoder for each frame. The highest value within the DC coefficients band, however, is assumed fixed. The process of reconstruction for each DCT coefficients band starts by comparing the quantized value of the DCT coefficient and the index of quantization interval that the corresponding side information DCT coefficient value falls in. This comparison can yield one of three cases as follows

I. Case 1: When the index of the quantization interval that the side information DCT Y^{DCT} is equal to that of Turbo decoded quantized symbol q', then the reconstructed DCT coefficient, X_l^{DCT}, is considered equal to the side information DCT coefficient.

II. Case 2: When the index of the quantization interval that the side information DCT coefficient Y_l^{DCT} is lower than that of the Turbo decoded quantized symbol $\acute{q}$ then the reconstructed DCT coefficient, $\acute{X}_l^{DCT}$, assumes the lower bound of the quantization interval indexed by the decoded quantized symbol.

III. Case 3: When the index of the quantization interval that the side information DCT coefficient Y_l^{DCT} is higher than that of the Turbo decoded quantized symbol $\acute{q}$ then the reconstructed DCT coefficient, $\acute{X}_l^{DCT}$, assumes the upper bound of the quantization interval indexed by the decoded quantized symbol.

The DCT coefficients bands for which no parity bits are sent are replaced by the corresponding DCT bands of the side information Y_l^{DCT} .

3.5 Chapter Summary

In this chapter the proposed Region-Based Adaptive DVC codec solution is introduced. The proposed solutions are characterized by adaptive partitioning process that considers the local characteristics of the video sequence and results in different coding units/region, where each coding units/region having appropriate coding parameters. The high level abstraction architecture of the proposed solution encloses three main components, the frame partitioning and parameters estimation, WZ-encoder and WZ-decoder. The proposed solution in this chapter is realized in the bi-directional communication scenario where the feedback channel is available and implemented in two domains, the pixel domain and the transform domain. The development of the theoretical foundations for implementing the solutions in each scenario is explained.

CHAPTER 4
REGION-BASED DVC CODEC WITHOUT FEEDBACK IN PIXEL DOMAIN

A part of this work is dedicated to address the rate control technique in the feedback based DVC system with the main objective to remove the feedback channel and propose DVC system that is applicable in the unidirectional scenarios. In section 4.1 of this chapter a brief introduction to an encoder rate control algorithm based on a machine learning algorithm is presented. A simplified walkthrough the proposed feedback free architecture is presented in section 4.2. Section 4.3 presents the frame partitioning and DVC parameter estimation. The rate estimation process is presented in section 4.3.3. Section 4.3.3.1 explains the machine learning technique used in the proposed solution. The proposed solution uses rate estimation module which is set up and prepared offline where the data for training the machine learning is collected and the machine is trained as explained in section 4.3.3.2. Section 4.3.3.4 explains how the trained machine is used online during the encoding process after the machine set up is completed. Section 4.4 presents the encoder of the proposed solution. Section 4.5 presents the decoder. Section 4.6 summaries the chapter.

4.1 Proposed Region-Based DVC Solution without Feedback in Pixel Domain

This section introduces the proposed Region-Based DVC without feedback solution in pixel domain. The Region-Based concept is employed in the proposed solution to improve the overall performance of the proposed solution motivated by the Region-Based coding advantages explained in chapter 3. Since the feedback is unavailable the encoder has to develop its own side information to perform the frame partitioning. To keep the encoder as simple as possible, a rough estimate of the side information is generated by averaging the two consecutive key frames. The frame partitioning is based on the SAD criterion, where the SAD is computed using the current WZ frame and the co-located blocks in the roughly estimated side information. The block with similar SAD are merged together to construct a region. To keep the encoder simple, the number of regions is predetermined experimentally by using an offline experiment. The rate estimation is performed at the bitplane level of the region in two steps. First, the encoder uses the side information of a given region to calculate the initial rate $H(X|Y_{enc})$. As the side information at the decoder and the encoder differ from each other, the encoder must estimate an extra rate ΔR to compensate for it. The extra rate is estimated using a simple machine learning algorithm, namely, k-Nearest-Neighbors, [83-87] to adhere to the low-encoding complexity constraint. The algorithm classifies the extra rate into one of 15 rate classes. The classification is based on the Euclidian distance between some attributes that characterize the quality of side information and the representatives of the rate classes. The rate classes' representatives are obtained during an offline training technique using a comprehensive data set. If the training data is very comprehensive and contains all possible ranges of input attributes then it can perform very well for different video sequences and therefore reduce complexity and delay.

4.2 A The Proposed Region-Based Adaptive Codec Solution without Feedback Architecture

The proposed Region-Based Adaptive without Feedback Architecture encloses the three main parts of the higher level abstraction in Figure 3-1, zooming into these three parts gives a detailed view of their components, as follows

Encoder:

- Conventional coding.
- Source coding (Scalar Quantizer).
- Slepian-Wolf Coding.

Frame partitioning and DVC parameter estimation

- Frame partitioning and DVC parameters estimation (Rate)

Decoder

- Conventional decoding
- Frame partitioning
- Side information generation
- Slepian-Wolf decoding.
- Reconstruction

Region-Based Adaptive DVC codec without feedback depicted in Figure 4-1, includes new rate control module. This module is set up offline before the system deployed to perform the video coding process. Therefore, the walkthrough will start by explaining through the set up of the rate control module.

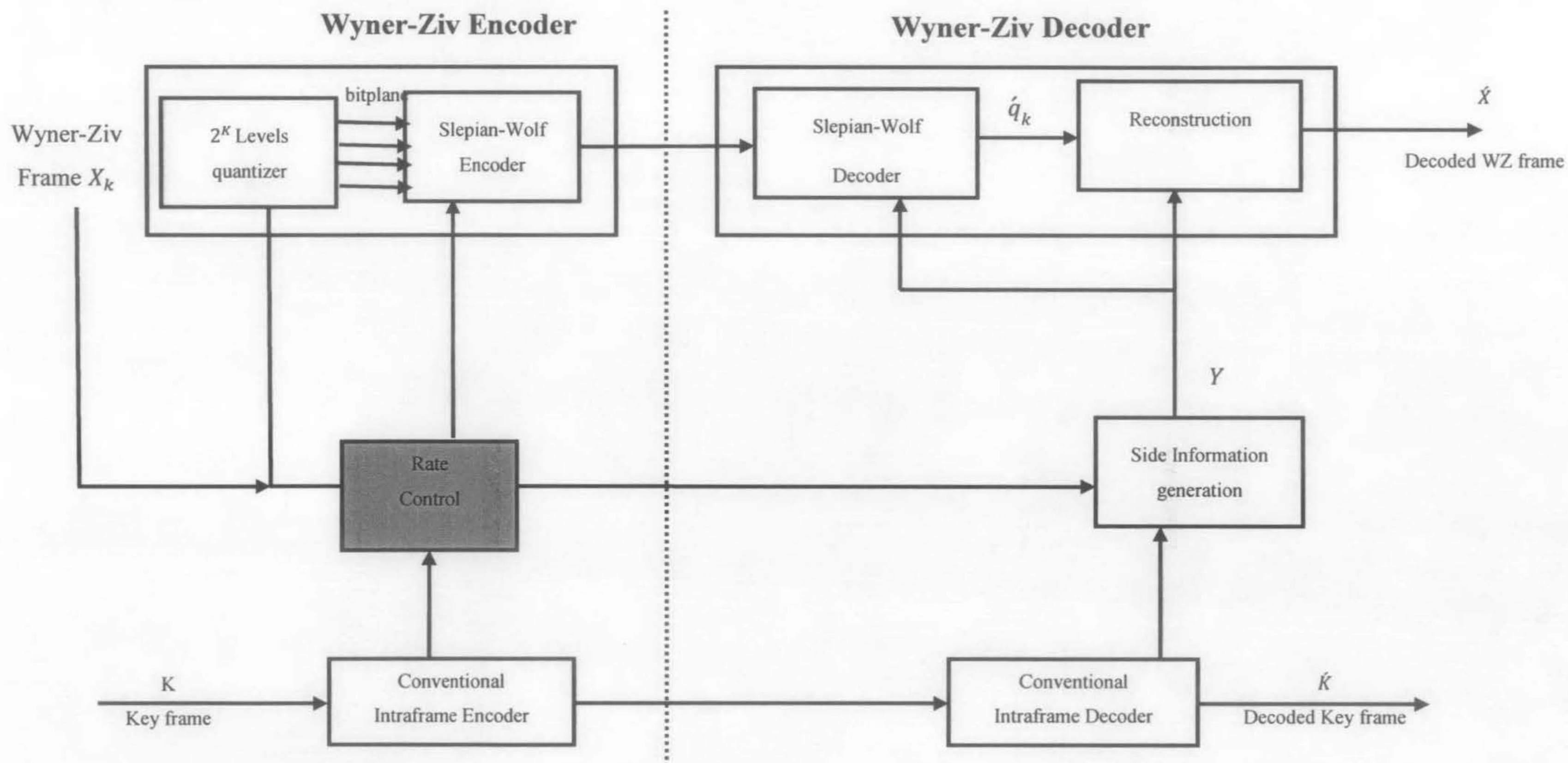

Figure 4-1: Feedback Free proposed architecture

4.2.1 Simplified Walkthrough of the Rate Control Module Set Up Procedure

- The rate is represented by classes that cover all possible number of requests.
- Training data is collected assuming the feedback is present.
- The included machine learning is trained using the training data mentioned above and the associated feature (attributes) for each rate class identified.

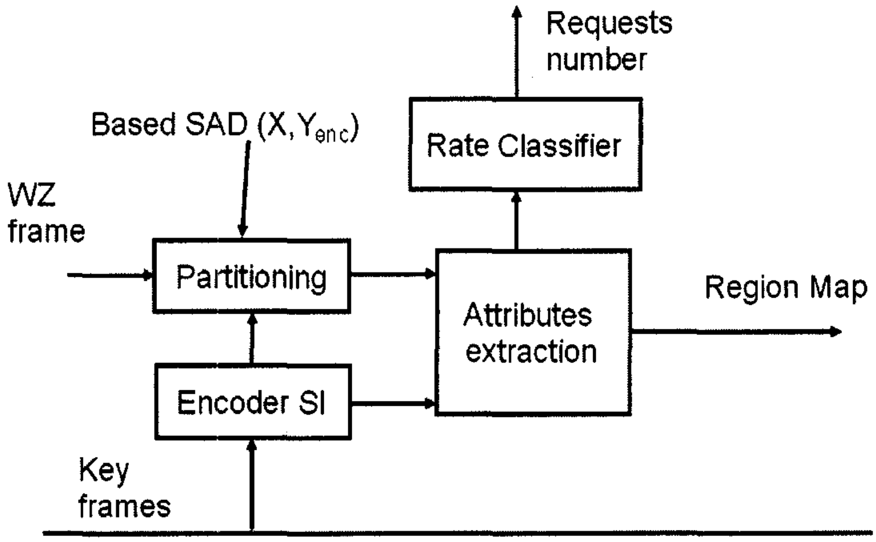

Figure 4-2: Rate estimation module at encoder

4.2.2 Simplified Walkthrough of Region-Based DVC Solution in Pixel Domain

1. The key frames (odd frames) are conventionally encoded and sent to the decoder.
2. The encoder creates its own version of the side information, by averaging the two consecutive key frames.
3. The partitioning process is performed based on the SAD difference between the WZ blocks and the co-located blocks in the encoder version of the side information
4. The WZ frame and encoder version of the side information are partitioned into L regions $\{X_1^{dec}, X_2^{dec}, \ldots . X_L^{dec}\}$ and $\{Y_1^{dec}, Y_2^{dec}, \ldots . Y_L^{dec}\}$
5. The corresponding regions map is compressed by using a quad tree algorithm and transmitted to the decoder.
6. The decoder decomposes the compressed regions map and partitions the side information into corresponding L $\{Y_1^{dec}, Y_2^{dec}, \ldots . Y_L^{dec}\}$ regions.

7. The samples of the first region X_l are quantized, generating the quantized symbol stream.
8. The Slepian-Wolf encoder produces redundant information (e.g. parity bits) from the quantized symbol stream bitplanes starting with the MSB bitplane.
9. The encoder estimates the initial rate for the region starting with the MSB bitplane R_{ini}.
10. The encoder computes the related attributes (features) and triggers the machine-learning based classifier to estimate the extra rate class ΔR.
11. The encoder sends the parity bits packets corresponding to this estimated rate.
12. At the decoder, the quantized symbol stream is decoded through joint source-channel decoding with the aid of the region's side information Y_1^{dec}.
13. The decoded quantized symbol stream and the side information, Y_1^{dec}, are then used together in a reconstruction module to reconstruct the sequence X_l resulting in the estimates, $\hat{X}_l$. The region's Y_1^{dec}, typically a decoder's estimate of the region X_l to be WZ encoded is, however, available at the decoder.

4.3 Frame Partitioning and DVC Parameters Estimation

This solution is characterized by the absence of the feedback channel. Therefore, it differs from the feedback based rate control solution in where the DVC parameter estimation and that the frame partitioning is performed. The other components at the system such as the encoding and decoding part are the same as explained in section 3.3. Therefore the following sections will only explain the frame partitioning and the DVC parameter estimation.

4.3.1 Frame Partitioning

In the absence of feedback channel, the encoder must partition the frame on its own account. Hence, the encoder must generate its own version of side information (SI) Y_{enc}. The process of the side information generation at the encoders must be kept as simple as possible; which results in a rough estimation to side information. An averaging of the two neighboring key frames seems an appropriate method. Thus,

$$Y_{2i}^{enc} = \frac{X_{2i-1} + X_{2i+1}}{2} \tag{4-1}$$

where $Y_{2i}^{enc}$is the estimate of $2i^{th}$ WZ frame, called SI here. The encoder uses the original WZ frame and its side information generated by using (4-1) to perform the frame partitioning as explained in section 3.3.1.1. The regions map is then obtained and its representation is compressed using the quadtree representation (see section 3.3.1.1). However, unlike the proposed DVC with feedback channel, the region map in this case is transmitted to the decoder, so the decoder can use the same partitioning information to decode.

4.3.2 DVC Parameters Estimation

There are only two parameters that need to be estimated in this case, namely the dependency channel parameter, α , and the rate for each region. To maintain the simplicity of the encoder, the process of estimating the quantizer set selection is removed. Instead a simple quantization scheme is identified and associated with respective rate-classes during the offline learning process. The dependency channel parameter is only needed at the decoder to perform the decoding and the reconstruction (see Section 3.3.1.2). In the following subsection the rate estimation is explained in detail.

4.3.3 Rate Estimation

The encoder based on its version of the side information (eq.(4-1)) Estimates the number of requests per bitplane and sends the corresponding number of parity bits to the decoder. The process of estimating the rate given by the entropy of X conditioned on Y^{dec}, $H(X|Y^{dec})$, is described in section 3.3.1.3. Following the same logic, the rate can also be estimated at the encoder using the entropy of X conditioned on Y^{enc}. However, the required rate to successfully decode the decoder version of the side information is different from that estimated at encoder due to the following:

1. Rate estimation at the encoder is carried out using original key frames that are available at the encoder, while the estimation at the decoder is based on side

information. Therefore the rate calculated at the encoder is less as compared to rate calculated at the decoder.

2. Even when the decoder side information is of good quality, it will still require more bits than that calculated at the encoder since theoretical rate depends on position of error in the code. Therefore, to correct various codes with same number of errors, occurring at different positions in the code, one requires different number of parity bits.

Based on these observations, it can be said that the decoder requires an extra amount of parity bits $\Delta \boldsymbol{R}$ to successfully decode. The rate estimation process can thus be mathematically described by (4-2)

$$\boldsymbol{R_k = R_{ini} + \Delta R} \tag{4-2}$$

$$R_{ini} = H(X|Y_{2i}^{enc}) \tag{4-3}$$

To correctly estimate the value of ΔR in (4-2), the encoder must either estimate the decoder version of side information, or know the position of the errors in the side information at the decoder, which is a more difficult task to do at the encoder due to the requirement of simplicity. Therefore, the value of ΔR in (4-2) cannot be exactly calculated at the encoder. Instead, it can only be estimated. In such a scenario, machine learning techniques emerge as appropriate solution to this rate-estimation problem. This solution of the rate estimation problem involves offline learning process which helps in estimating extra rate, ΔR, for each rate-class and associates that with relevant features. This offline learning process is performed as though the feedback channel is present.

4.3.3.1 Machine Learning

Machine Learning is a branch of Computer Science that is concerned with designing systems that can *learn* from the provided input. Usually the systems are designed to use this learned knowledge to better process similar input in the future to make accurate predictions [84]. As depicted in Figure 4-3, the classifier is first fed *training data* in which each item is already *labeled* with the correct *label* or *class*. This data is used to train the *learning algorithm*, which creates *models* that can then be used to *label/classify* similar data. Since the overall target of the DVC is low-encoding complexity, the machine learning complexity and the new data classification must be kept as simple as possible. As the nearest-neighbors algorithm is considered one of the simplest machine learning algorithms [85] it proposed to the proposed rate estimation solution.

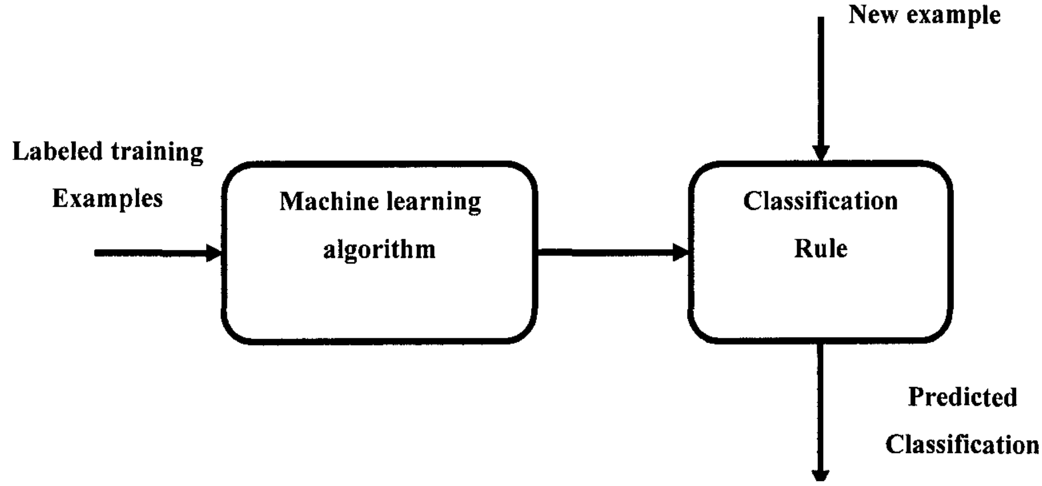

Figure 4-3 Basic Machine learning architecture

The k-nearest neighbor algorithm ($k - NN$) is a method for classifying objects based on the closest training examples. Objects are defined in terms of their features and the classification done in this feature space. $k - NN$ is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. With k-nearest neighbor algorithm an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of its nearest neighbor.

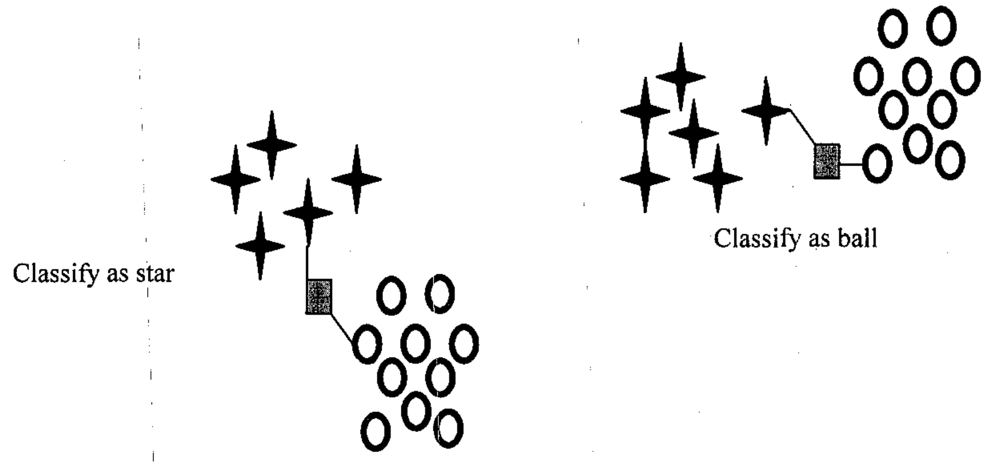

Figure 4-4 Nearest Neighbors demonstration

A modified version of the $k-NN$ algorithm has been introduced in [86] to reduce the classification time for $k-NN$ where the space is clustered (training data) in to subclasses. Each subclass will be represented by one data (two or more representatives according to the number of subclasses is possible). After that the classification process is performed by *NN (or* $-k-NN$*)* algorithm to classify by using the representative data. Each subclass contains a random number of data, which are relatively close to each other. The classification time in NN and $k-NN$ are of the order N^2, i.e., $O(N^2)$, where N is the training data size. In fact, the time of the classification will depend just on the number of subclasses cc_i, where cc_i the number of subclasses in the class C_i. In general μ_i is a small number and doesn't depend on the training data size. The $k-NN$ classifier is generally based on the Euclidean distance between a test sample x and the specified training samples.

4.3.3.2 Setting Up Phase: Rate Classes And Feature Association

The first step of setting up the classifier is the training process. This is done by feeding the classifier with *training data* in which each set of attributes (features) is already *labeled* with the correct *label* or *class*. The set of attributes (features) is a group of parameters that are associated with each label or class. To put these into the rate estimation context, the rate is represented by the number of requests. Feedback based DVC systems make use of a rate-compatible-punctured codes (RCPT) in their module of the SW codec [65] and the puncture window size is usually 8 [66], resulting in 16 puncture code rates. These are referred to as rate control classes. At the time of collecting training data, the feedback channel is assumed present that enables collecting training data to simulate the behavior of the feedback channel.

Several works have been carried out that analyze the impact of various attributes (metrics) on the feedback channel, e.g., the number of times the feedback channel is used (number of requests) as well as its associated rate [56, 87]. It is also desirable to study the improvement of the quality of the decoded frames as more parity bits are requested via feedback channel. This work has reported the following observations:-

1. The number of requests for bits increases with the number of bitplanes to be decoded.
2. The increase is not constant across the bitplanes. It depends mainly on the correlation between the side information and the WZ frame for each bitplane.
3. This correlation is usually high for the bitplanes with more significance as compared to the bitplanes with less significance.
4. It has also been observed that the reconstructed frame quality increases quickly with the number of requests until the average number of decoder requests is reached.
5. The average number of feedback requests for decoding a frame already implies that most of the frames are correctly decoded.
6. Increasing the number of requests beyond the average value comes with only a small increase in the average PSNR.

The main focus at this stage is to identify the attributes that describe the relation between the original region and the estimated region at the encoder to decide what attributes can describe the behavior of the feedback channel. Those attributes are important since they allow characterizing the usage of the feedback channel. Following is the list of attributes that have a potential to influence the estimation of rate at the encoder, some of which may influence it more significantly than others:

1) The error probability: The error probability P_e of the bit value b_k of pixel x is estimated using (3-12) where $P(X|Y_{2i}^{enc})$ is modeled by the Laplacian distribution.

2) The mean squared error (MSE): The MSE between the $X_l$of the current WZ region and Y^{enc} of side information is obtained by

$$\sigma^2 = \frac{1}{O} \sum_{i,j \in X_l} (X_l(i,j) - Y_l^{enc}(i,j))^2 \tag{4-4}$$

Here, Y^{enc} is updated by using the previous bitplanes in performing the reconstruction process.

3) The sum of absolute Difference (SAD): The SAD between the X_l of the current WZ region and Y^{enc} is obtained by

$$SAD = \frac{1}{O} \sum_{i,j \in X_l} (X_l(i,j) - Y_l^{enc}(i,j)) \tag{4-5}$$

The Y^{enc} is updated by using the previous bitplanes in performing the reconstruction process.

4) The Entropy of the difference: This entropy between the X_l of the current WZ region and Y^{enc} is obtained by

$$H(X_l - Y_l^{enc}) = \sum_{i \in SAD} (X_l(i) - Y_l^{enc}(i)) \log (X_l(i) - Y_l^{enc}(i)) \tag{4-6}$$

5) The initial estimate of number of request Q_{ini}: The estimate of Q_{ini} based on the side information at the encoder is computed by using (3-14)

$Q_{ini=}\frac{R_{ini,k}}{\rho}$ where ρ is packet size and $R_{ini,k}$ is initial rate for bit-plane k.

$\rho = \frac{O}{\omega}$ where O is the bitplane length and ω is the puncturing period.

6) The Region class: The region class is determined by the quality of the side information that helps partitioning.
7) The bitplane length: This is determined by the size of the region.
8) The target attribute that label each set of attributes is ΔR which is described as the number of added requests to perform the decoding process successfully

$$\Delta Q = \frac{\Delta R}{\rho} \tag{4-7}$$

4.3.3.3 The Criteria for Choosing Significant Attributes

The accuracy of the classifications process depends on the choice of the attributes used. Irrelevant attributes will confuse and therefore degrade the classifier performance. Therefore, all attributes that do not contribute to rate classification are removed and the most appropriate attributes required for classification are considered only.

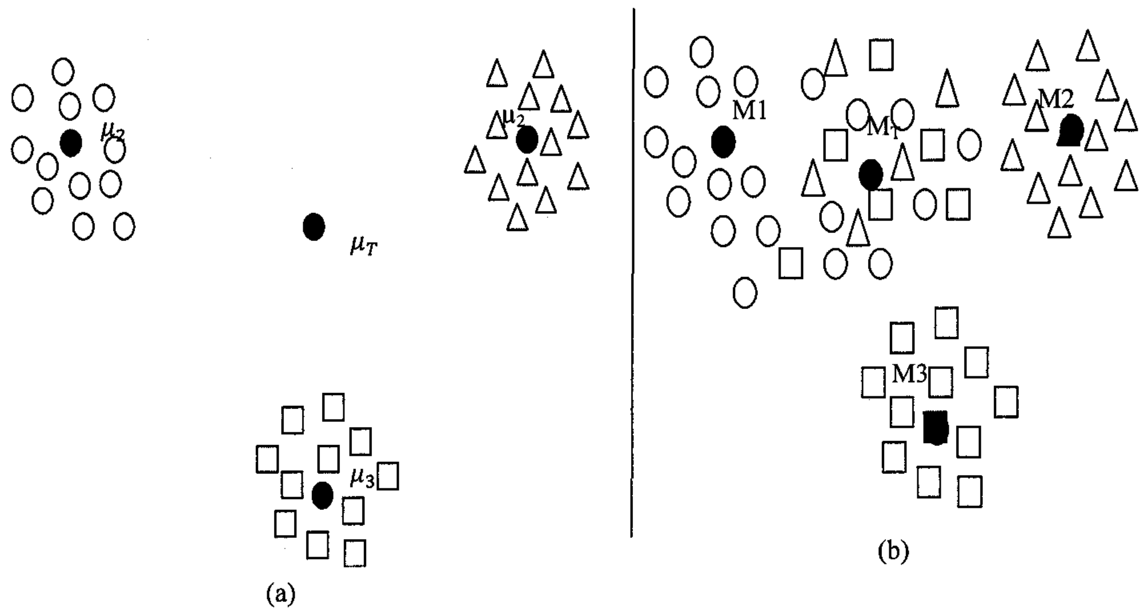

Figure 4-5: Distribution of attributes a) distribution of attributes A, b) distribution of attribute B

Assume 3 classes (1, 2, 3) and two attributes (A, B) as shown in Figure 4-5(a, b). Suppose μ_1, μ_2, and μ_3 are the means of class1, class2, and class3 respectively and the m_T is the mean total of all classes. The Figure 4-5 a) shows the distribution of the attribute A among the 3 classes. It shows that attribute A is a good metric for classification process and leads to better classification accuracy. Similarly, Figure 4-5-b shows the distribution of the attribute B among the classes. However, it can be seen from Figure 4-5-b that attribute B gives very confusing clues to the classification process due to the overlapping between the classes which leads to more error probability.

Assume $A = \{A_{11}, A_{12}, A_{13}, \ldots . A_{21}, A_{22}, A_{23}, \ldots . A_{31}, A_{32} A_{33}\}$ and $B = \{B_{11}, B_{12}, B_{13}, . B_{21}, B_{22}, B_{23}, . B_{31}, B_{32} B_{33}\}$ Where$\{A_{11}, A_{12}, A_{13}, , , , \}$ and

$\{B_{11}, B_{12}, B_{13}, \dots\}$ instances from attribute A and B for class 1. For each class we obtain the variance Var_i and the mean μ_i we obtain mean of the means μ_T

$$\mu_T = \frac{1}{C}\sum_{i=1}^{C}\mu_i \tag{4-8}$$

The modified variance is obtained by

$$\sigma = \frac{1}{C}\sum_{i=1}^{C}\frac{(\mu_i - \mu_T)^2}{var_i} \tag{4-9}$$

The attribute with largest value for σ contributes the most in the classification process. To avoid the attributes that complicates the classification process for each bitplane different but significant attributes are used based on the value of σ.

Since the behavior of the feedback channel varies from one bitplane to another, performing this test on each bitplane attributes can result in different number of significant attributes to be used as relevant feature for rate classes within each bitplane.

This criteria has been used to choose the significant attributes for each bitplane and it was found that the error probability P_e and the MSE have a highest variance for all bitplanes. The significance of the remaining attributes varies from one bitplane to another. For example, for the 3rd bitplane, only four attributes show reasonable values of variances (P_e , MSE, SAD, Entropy), while for the 4th bitplane, all attributes are of significance and therefore chosen.

4.3.3.4 Using the Proposed Codec

Once the training is complete and each class rate is associated with a set of appropriate attributes. The built classifier (rate estimation module) is attached to the system as shown in Figure 4-1 and the system can work without the need for feedback channel. The process of the attached classifier (rate estimation module) starts by collecting the corresponding attributes presented in section (4.3.3.2) for the bitplane under process. The appropriate attributes for each bitplane are selected based on the outcome of offline performed significance criteria test (see section 4.3.3.3). Given a set of training cases/objects and the selected attribute values, the classifier predicts the target attribute ΔQ the number of requests to be added to the initial number of requests Q_{ini}. The total number of requests is given by.

$$Q_{total} = Q_{ini} + \Delta Q \quad (4\text{-}10)$$

The estimated rate is given by

$$R_k = \rho \times Q_{total} \quad (4\text{-}11)$$

4.4 WZ-Encoder in the Proposed Solution

Although the setting up of the rate estimation module, namely, machine learning based classifier is performed offline (4.3.3.2) and attached to the proposed solution WZ-encoder, the WZ-encoder needs to collect set of attributes as explained in section 4.3.3.2 and perform the rate estimation process (4.3.3.4). The theoretical foundation for the pixel domain source coding is explained in 3.3.2.1. The WZ-encoder performs source coding (quantization). The resulting symbols of the quantized pixel are converted into bitplanes array and fed into the Slepian-Wolf encoding. The Slepian-Wolf encoding as explained in section 3.3.2.3 generates two types of information, namely, the systematic information bits and the parity bits. The systematic information is discarded and the rate estimation module predicts the number of requests and accordingly the number of parity bits to be sent per bit-plane of a given region.

4.5 WZ-Decoder in the Proposed Solution

The decoder in the feedback based systems performs rate control process. However, since the proposed solution is feedback free, the decoder does not participate in the rate control process anymore. The decoder, in this case, only requires decomposing the received region's map to partition the side information. The remaining tasks of the decoder used in feedback based systems are retained in the proposed system such as side information generation and Slepian-Wolf decoding and the reconstruction. The decoder generates the side information as explained in 3.3.3.1 and uses the received partition map to partition this side information into corresponding regions. The received parity bits and the side information are used in the Slepian-Wolf decoding process as explained in 3.3.3.2. In case the received parity bits amount is not sufficient for successful decoding the decoder current bitplane is replaced by the decoder's estimated bitplane. The last process performed by the WZ-decoder is the reconstruction process and it is performed as explained in section 3.3.3.3.

4.6 Chapter Summary

In this chapter pixel domain Region-Based Adaptive DVC codec without feedback channel is introduced. The proposed solution is characterized by encoder rate estimation module that enables the DVC codec to work under uni-directional channel constraints. This scenario differs from the Region-Based Adaptive DVC codec in pixel domain in where and how to perform frame partitioning and parameters estimation while preserving the simplicity of the WZ-encoder. The partitioning process is performed at the encoder which requires the encoder to create its own side information and transmits the partition/region map to the decoder side. Although communicating the region map to the decoder adds rate penalty, but the overall rate-distortion performance of the system is expected to improve by employing the region based concept in the proposed solution. The high level abstraction architecture of the proposed solution encloses three main components, the frame partitioning and parameters estimation, WZ-encoder and WZ-decoder. The encoder complexity is increased by the attached rate estimation module but considering the traditional video coding encoder the proposed solution is simpler since the motion estimation/compensation process is not performed. The decoder performs the Slepian-Wolf decoding once for each bitplane since no request can be made, which reduce the time and the complexity of the proposed solution WZ-decoder.

CHAPTER 5

METHODOLOGY OF PERFORMANCE EVALUATION

In this chapter, the methodology of evaluating the proposed solutions presented in chapter 3 will be covered in more detail. Section 5.1 presents the performance evaluation metrics used to evaluate the proposed solutions. Section 5.2, presents the video sequences used as test materials. Section 5.3 presents the evaluation methodology and its associated parameters of the Region-Based Adaptive DVC codec in pixel domain. Section 5.4 covers the same for Region-Based Adaptive DVC codec evaluation in the transform domain. Section 5.5 the parameters and the methodology for Region-Based Adaptive DVC codec without feedback solution is presented.

5.1 The Performance Evaluation Metrics

5.1.1 Rate-Distortion Curve

To evaluate the proposed RD performance of the DVC codec solutions, the rate-distortion plots only contain the rate and the Peak Signal-to-Noise ratio (PSNR) values for the even frames, since the GOP is fixed to be 2. To evaluate the video codec performance, different rate points are used - each rate point corresponds to a particular quantizer. The rate is the number of bits per second transmitted to decode the encoded frame. In the DVC codec, the rate is the number of transmitted parity bits per second. In the proposed Region-Based Adaptive DVC codec, extra bits are used to communicate the regions map. The (PSNR) is the ratio of the peak signal to the root mean square (RMS) noise, where the noise is the difference between the reconstructed frame after encoded by the video codec under test and a reference copy of the same original frame PSNR objectively measures degradation primarily of the luminance signal (5-1). Acceptable PSNR varies between 20 dB (maybe just acceptable) and 40 dB (excellent) [88].

$$MSE = \frac{1}{XY}\sum_{x=1}^{x=X}\sum_{y=1}^{y=Y}(original(x,y) - Reconstructed(x,y))^2 \qquad (5\text{-}1)$$

Where X and Y are the frame width and height respectively. MSE is converted to peak signal-to-noise ratio (PSNR), in decibel (dB), where PSNR is defined by (5-2) as:

$$PSNR = 20\ \log_{10}\frac{255}{MSE} \qquad (5\text{-}2)$$

5.1.2 Complexity

The complexity of the video coding system encoder represents a great challenge for video applications with limited resource. The proposed solution adheres to the requirements of the DVC codec by shifting the motion estimation/compensation towards the decoder side. Its encoder differs from the *state-of-the-art* Frame-Based DVC system encoder by performing the WZ frame partitioning at the encoder. But this process is minimized to its minimum by estimating the partition map at the decoder and the decoder only needs to interpret the partition map and partition the WZ frame (see section 3.3.2.1) on the received partition map. This extra complexity helps the DVC codec to benefit from the Region-Based video coding advantages and does not contradict with DVC encoding requirements.

5.2 Test Material

Four video test sequences have been selected for the evaluation of performance of the proposed Region-Based Adaptive DVC codec solutions proposed in this thesis. A brief description of the main characteristic of each test sequence is provided in Table 5-1, (fps stands for frames *per* second).

In Table 5-1 Test Material, the “Miss America” and “News” test sequences are examples of video conference content typically characterized by a low and medium low amount of movement (activity). The “Mother-Daughter” and “Carphone” test sequences are examples of video content characterized by medium higher and high activity. As it is well-known, the higher the activity, the more difficult it is to compress the video content. This content variety is important to collect enough representative and meaningful results for the proposed Region-Based Adaptive DVC codec performance.

Table 5-1 Test Material

Video Sequence Name	*Carphone*	*Mother and daughter*	*Miss America*	*News*
Sample Frame				
Total Number of Frames	382	961	150	300
Number of Used Frames	101	101	101	101
Spatial Resolution	QCIF	QCIF	QCIF	QCIF
Temporal Resolution (fps)	30	30	30	30

5.3 Region-Based Adaptive DVC Codec Pixel Domain Test Methodology

In this section, the methodology for the proposed Region-Based Adaptive DVC codec pixel domain performance evaluation is described:

- Only the luminance data is considered in the Region-Based Adaptive DVC codec pixel domain (PD) rate-distortion performance evaluation.

- The five quantizers are used to obtain 4 different RD points for the Region-Based Adaptive DVC codec PD solution [1, 2, 3, 4,5].
- The key frames, shown in Figure 3-2, are intracoded by MPEG+.
- In the turbo encoder implementation, two recursive systematic convolutional encoders of rate ½ are employed; each one is represented by the generator matrix $\left[1 \quad \frac{1+D+D^2+D^4}{1+D^2+D^4}\right]$
- The side information is generated by using the algorithm presented in section 3.3.2.
- To compare the performance of the Region-Based Adaptive DVC codec PD solution with the Frame-Based PD, each test sequence is encoded by both the Region-Based Adaptive DVC codec PD and the Frame-Based PD.
- For the Region-Based Adaptive DVC codec PD case; as in [2, 65], a Laplacian distribution is used to model the residual between the X_l region pixel values and the corresponding pixel values of the Y_l region frame.
- For the Frame-Based PD case; as in [11, 65], a Laplacian distribution models the residual between the X_{2i} frame pixel values and the corresponding pixel values of the Y_{2i} frame.
- The Laplacian parameter is adaptively estimated by using the online channel model parameters estimation as presented in [57]; for both cases the Frame-Based and the Region-Based Adaptive DVC codec PD.
- The maximum allowable number of turbo decoding iterations is 20[65].
- The bit error rate threshold is assumed to be 1×10^{-3}; as it is commonly used in relevant work [2, 64, 65].
- To evaluate the Region-Based Adaptive DVC codec PD codec RD performance, the rate-distortion plots only contain the rate and the PSNR values for the even frames, i.e. the WZ coded frames, of a given video sequence.
- For each test sequence, the Region-Based Adaptive DVC codec PD RD performance is compared against MPEG+ interframe coding with I-P-I-P structure. In the later case, only the rate and PSNR of the P frames is shown since the Region-Based Adaptive DVC codec PD DVC codec performance is the target.

5.4 Region-Based Adaptive DVC Codec Transform Domain Test Methodology

In this section the methodology for the Region-Based Adaptive DVC codec transform domain performance evaluation is presented

- Only the luminance data is considered in the Region-Based Adaptive DVC codec transform domain (TD) rate-distortion performance evaluation.
- The key frames, presented in figure 3-3, are intracoded by MPEG2.
- In the turbo encoder implementation, two recursive systematic Convolutional encoders of the rate ½ are employed; each one is represented by the generator matrix $\left[1 \;\; \frac{1+D+D^2+D^4}{1+D^2+D^4}\right]$
- The side information is generated by using the algorithm presented in section 3.3.2.
- The number of partitions is predefined for each test sequence based on the motion activity within the test sequence.
- Each region is divided into 4×4 blocks, these transformed into the DCT transform domain. The DCT coefficients are grouped together according to the position occupied by the DCT coefficients within the 4×4 DCT coefficients blocks, forming the DCT coefficients bands.
- For the X_l region, the dynamic range of each DCT band is assumed to be losslessly available at the decoder.
- The 7 quantization matrices depicted in Figure 3-10 are used to obtain 8 different RD points for the Region-Based Adaptive DVC codec TD solution.
- To compare the performance of the Region-Based Adaptive DVC codec TD solution with the Frame-Based TD each test sequence is encoded by both the Region-Based Adaptive DVC codec TD and the Frame-Based TD.
- For the Frame-Based TD case; as in [13, 65], a Laplacian distribution is used to model the residual between the X_{2i} frame DCT coefficients and the corresponding DCT coefficients of the of the Y_{2i} frame.
- For the Region-Based Adaptive DVC codec TD case; as in [2, 65] a Laplacian distribution is used to model the residual between the X_l region DCT coefficients and the corresponding DCT coefficients of the Y_l region.

- The Laplacian parameter is adaptively estimated by using the online channel model parameter estimation as presented in [57]; for both cases the Frame-Based TD and the Region-Based Adaptive DVC codec TD.
- The maximum allowable number of turbo decoding iterations is 20.
- The bit error rate threshold is assumed to be 1×10^{-3}; as it is commonly used in relevant works [13, 64, 65].
- Since what is important at this stage is to evaluate the Wyner-Ziv codec RD performance, the rate-distortion plots only contain the rate and the PSNR values for the even frames.
- For each test sequence, the Region-Based Adaptive DVC codec TD RD performance is compared against the MPEG+ interframe coding with an I-P-I-P structure. In the later case, only the rate and PSNR of the B frames are shown since the Region-Based Adaptive DVC codec TD DVC codec performance is the target here.

5.5 Region-Based Adaptive without Feedback DVC Codec Methodology

5.5.1 The Set up Methodology (Offline)

The first step in using the classifier as explained in chapter 4, is the training process to build the classifier. To perform the training with comprehensive data 4 different video sequences were used "Hall Monitor", "Foreman", "Car phone" and "Coastguard". To gain some knowledge regarding the behavior of the feedback channel, the histogram of the rate classes for different bitplanes is plotted. From Figure 5-1 we can see that the higher classes (10-15) have very low frequency. Since the optimization of the training process depends on the amount of the training data set, the higher classes (10-15) are merged together in one class (13).

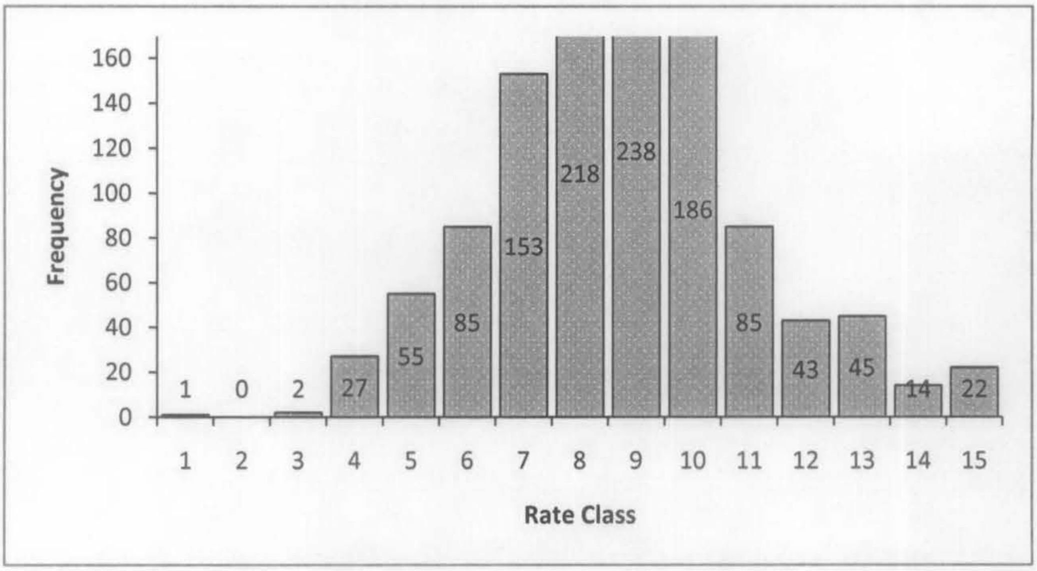

Figure 5-1 The histogram of the rate classes for the first bitplane (QP = 4).

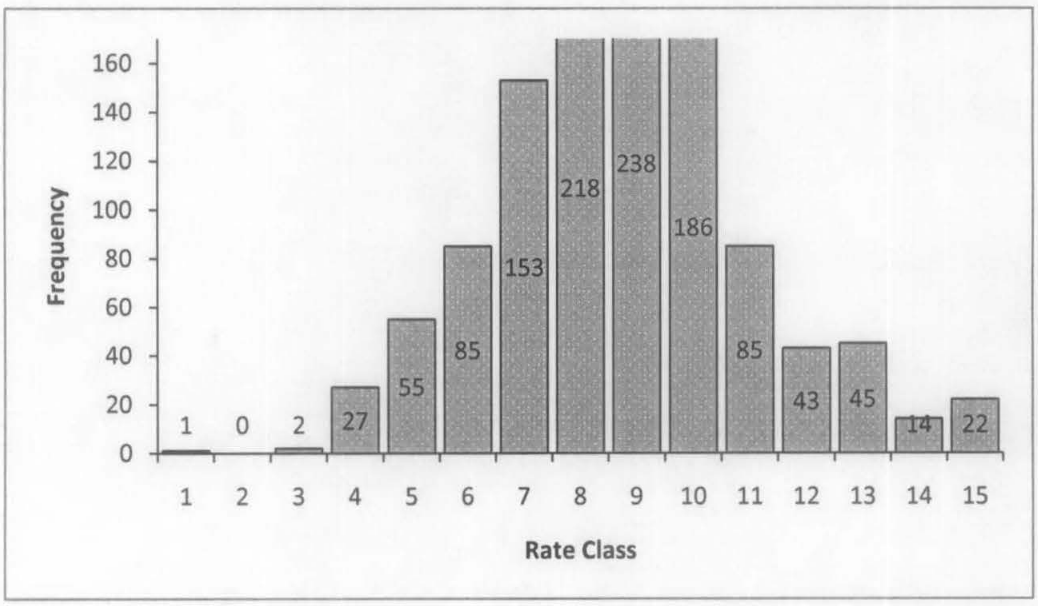

Figure 5-2 The histogram of the rate classes for the first bitplane (QP = 4).

From figure 5-2 we can see that the classes (8-12) have a very low frequency and the frequency of classes (13-16) is nearly zero. Therefore, the higher classes (8-16) are merged together in one class (9).

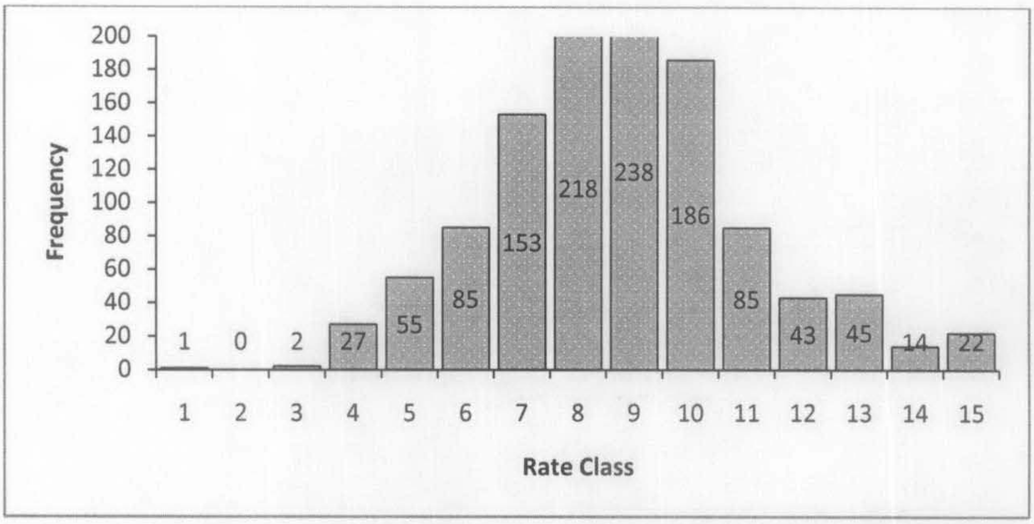

Figure 5-3: The histogram of the rate classes for the second bitplane (QP = 4)

From figure 5-3 we can see that the lower classes (1-4) and classes (10-12) have a very low frequency and the frequency of classes (13-15) is nearly zero. Therefore the lower classes (1-4) are merged together into one class (4) and the classes (10-12) are merged into a single class (11).

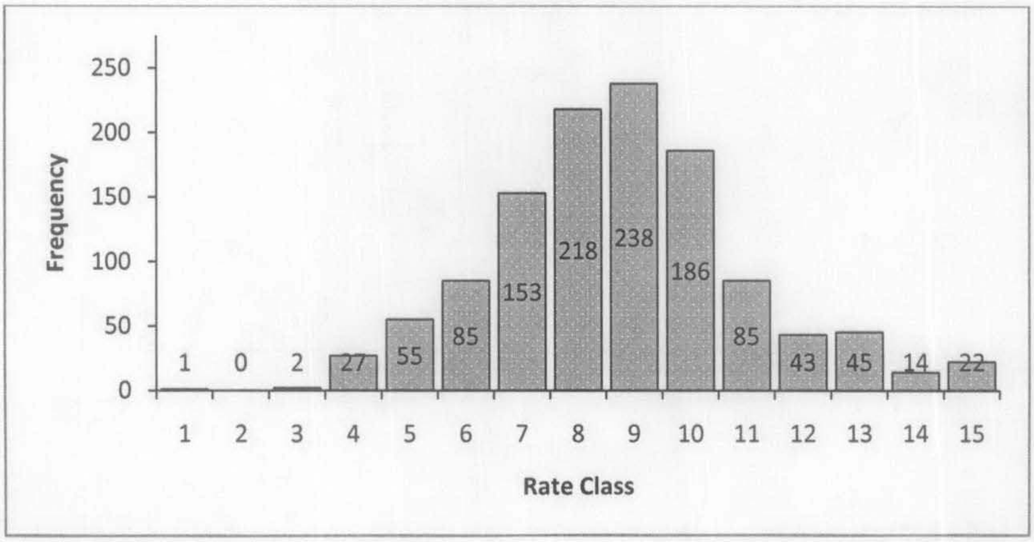

Figure 5-4 The histogram of the rate classes for the third bitplane (QP = 4).

From figure 5-4 we can see that the lower classes (1-4) and classes (11-15) have a very low frequency. Therefore, the lower classes (1-4) are merged together into one class (4) and the classes (11-12) are merged into a single class (13).

Table 5-2: The rate class's frequency per bitplane

Bitplane	Classes 1-15								
Bitplane1	140	140	164	168	135	126	101	79	60	55-9-5-1-0-1
Bitplane2	116	188	164	186	192	112	71	74	49-19-11-1-1-0-0	
Bitplane3	1-7-29-66	130	176	240	271	157	57	27-17-6-0-0		
Bitplane4	1-0-2-27-55	85	153	218	238	186	85	43	45-14-22	

Table 5-2 shows the frequency of the different class rate, for each bitplane. Since each bitplane has it is own characteristics, 4 classifiers will be built for each bitplane. To obtain sufficient data for training the classifiers, the rate classes with low frequencies are merged together and the merged group is given the higher class rate. The result of this merging is shown in Table 5-3, for bitplane 1 the classifier will classify the target rate into 1-10 classes and start from class 1 up to class 10. For bitplane 2 the classifier will classify the target rate into 1-9 classes starts from class 1 up to class 9, the bitplane classifier starts from class 4 and ends at class 12 with 9 possible classes, for the last bitplane the classifier starts from class 5 and ends at class 13 with 8 possible rate classes.

Table 5-3: Merged classes with low frequencies

bitplane	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Class9	Class10
Bitplane1	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
Bitplane2	R1	R2	R3	R4	R5	R6	R7	R8	R9	
Bitplane3	R4	R5	R6	R7	R8	R9	R10	R11	R12	
Bitplane4	R5	R6	R7	R8	R9	R10	R11	R12	R13	

As a consequence to the classes merging, it is expected to encounter some overestimation and underestimation cases. Although no improvement is gained in PSNR value in an overestimation case, this over rate is seen as an acceptable rate penalty. In the case of an underestimation, the decoding is deemed unsuccessful. Since the feedback channel is unavailable, the decoder uses the corresponding bitplane at the decoder. Using some of the decoder side bitplanes in the reconstruction degrades the quality of the reconstructed regions therefore, a proposed feedback free architecture has to be implemented such that the number of underestimations is decreased as close to zero as possible. Since the underestimation cases are undesired, the classification process must favor the overestimation option. The process is performed hierarchically where the classifier classifies the targeted rate class into one of two classes; the first class is the lowest rate class and the second class represents all the higher rate classes. This way the classification process eliminates the low rate class where it is more likely to experience underestimation cases. Figure 5-5 shows the classification process flow for the case of 10 rate classes. In the first step, the classifier classifies the new rate class into either class one (one parity bit packet) or into merged class (2-10), this class represents all rate class higher than class 1. If the classifier chooses merged class (2-10), another classification process is needed where the classifier chooses class 2 or merged class (3-10), the classification process is repeated until the classifier classifies the rate into a single class.

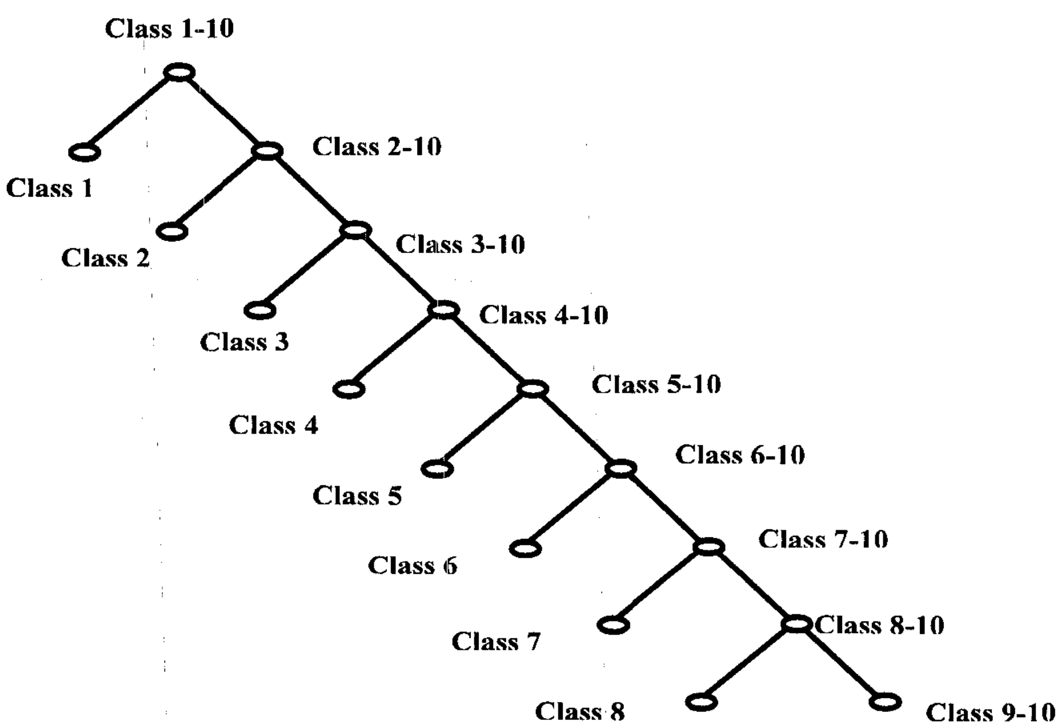

Figure 5-5 Hierarchy Classification

5.5.2 Rate Estimation (Online) Methodology

In this section the methodology for Region-Based Adaptive DVC codec without Feedback performance evaluation is explained.

- Only the luminance data is considered in the Region-Based Adaptive DVC codec without Feedback rate-distortion performance evaluation.
- The key frames, presented in Figure 4-1, are intracoded by MPEG2.
- The encoder creates its own version of the side information Y^{enc} by averaging the two consecutive key frames (X_{2i-1}andX_{2i+1}).
- The partitioning process is performed based on the SAD difference between the WZ blocks and the co-located blocks in the encoder version side information and WZ frame and encoder version of side information are partitioned into *4-5* L regions $X_1, X_2, \ldots. X_L$ and $\{Y_1^{dec}, Y_2^{dec}, \ldots. Y_L^{dec}\}$ the corresponding regions map is created by using quad tree algorithm.

- In the turbo encoder implementation, two recursive systematic Convolutional encoders of rate ½ are employed; each one is represented by the generator matrix $\begin{bmatrix} 1 & \frac{1+D+D^2+D^4}{1+D^2+D^4} \end{bmatrix}$
  - The 4 quantizers are used to obtain 4 different RD points for the Region-Based Adaptive DVC codec without Feedback solution [1, 2, 3, 4].
  - To compare the performance of the Region-Based codec without Feedback solution with the Region-Based PD each test sequence is encoded by both the Region-Based Adaptive DVC codec without Feedback and The Region-Based Pixel Domain.
  - For the Region-Based Adaptive DVC codec without Feedback case; as in [2, 65] a Laplacian distribution is used to model the residual between the X_l region pixel values and the corresponding pixel values of the Y_1^{dec} region.
  - For the Region-Based PD case; as in [8, 65], a Laplacian distribution is used to model the residual between the X_l region pixel values and the corresponding pixel values of the Y_1^{dec} region.
  - The Laplacian parameter is adaptively estimated by using the online channel model parameters estimation as presented in[57]; for both cases the Region-Based Adaptive DVC codec PD and Adaptive DVC codec without Feedback.
  - The maximum allowable number of turbo decoding iterations is 20.

- The bit error rate threshold is assumed to be 1×10^{-3}; as it is commonly used in relevant work [2, 64, 65].
- Since what is important at this stage is to evaluate the Region-Based Adaptive DVC codec without Feedback RD performance, the rate-distortion plots only contain the rate and the PSNR values for the even frames.

- For each test sequence, the Region-Based Adaptive DVC codec without Feedback RD performance is compared against MPEG+ interframe coding with an I-P-I-P structure. In the later case, only the rate and PSNR of the P frames is shown since the Region-Based Adaptive DVC codec Pixel Domain DVC codec performance is the target.

5.6 Chapter Summary

In this chapter, the methodology of evaluating the proposed solutions is presented. To assess the performance of the proposed solutions this chapter presented the test video sequences with different motion characteristics. The testing parameters and experimental setting for the proposed solutions is clearly stated so the reader can easily reproduce the results. The quantization and Slepian-Wolf codec for each scenario of the proposed solution are explained in this chapter. The validation is made comparing the results with those of popular frame-based DVC solution and MPEG+ codec. The difference between the implementation of the proposed solutions in pixel domain and transform domain can be readily observed in this chapter. The statistics analysis of the behavior of the feedback channel required in setting up the proposed feedback free solution is presented.

CHAPTER 6
RESULTS & DISCUSSION

This chapter presents the results of the performance evaluation of the proposed solutions in the chapter 3 and chapter 4 based on the methodology presented in chapter 5. Section 6.1 presents and discuses the performance of the frame partitioning process. Section 6.2 evaluates the performance of the Region-Based Adaptive DVC codec proposed in chapter 4 in which the proposed solution has two features, the accurate dependency channel model and the adaptive source coding. The section discuses how each one of these two features contributes in improving the performance of the proposed DVC codec comparing to the Frame-Based DVC codec. Section 6.3 presents and discuses the impact of the two features on implementing the proposed solution in the transform domain. Section 6.4 presents the results and the related discussion of the performance of the Region-Based Adaptive DVC codec without feedback channel in the pixel domain performed based on the methodology presented in chapter 5.

6.1 Analyses of Frame Partitioning

The partitioning process partitions the frame into partitions based on the local characteristics of the video sequence, as a result of which for example, the nonmoving background part is grouped into one region. Those parts of the frame with a residual higher than certain threshold can be grouped into one or more regions. For GOP =2, the scene change is not too much and the background part of the frame represents the largest part of the frame. For video sequences with low motion activity (Miss America) the background represents 80 % of the frame size and for the sequence with high motion activity (Carphone) the background represents around 60%. In both cases the foreground represents 20-40% of the frame size. This foreground contains several numbers of objects with different motion and, as mentioned in section 3.3.1.1, it can be divided into one or more regions. Partitioning the foreground into more regions results in regions of different sizes. For example, for low activity video sequence, partitioning the foreground (the 20% of the frame size) into 3 regions will produce region with average size around 7% of the frame size and for the high motion activity regions will produce region with average size around 14% of the frame size. As shown in Table 6-1 increasing the number of the regions decreases the size of the resulting regions, which subsequently decreases the length of the bitplanes.

Table 6-1 Resulting region size when partitioning the foreground

Number of regions in the foreground	Low motion activity Region size (Miss America, News)			High motion activity region size (Carphone & mother-daughter)		
	average Size percent	Bitplane length	Region map length	Size percent	Bitplane length	Region map length
3	6 -.05%	1500-1000	182	13-0.5%	3400-2500	202
4	5-0.01%	1100-850	230	10-0.3%	2540-2000	254
5	4-1.05%	900-650	450	8-0.2%	2000-1500	480
6	3-0.1%	750-500	635	6-0.1%	1500-1100	625
7	2.2-0.05%	500-340	750	5-0.02%	1100-900	740

The Slepian-Wolf codec in the proposed solution uses Turbo code, the performance of which depends heavily on the size of the block (bitplane). These codes perform poorer for shorter block size (bitplane). Another factor influencing the choice of number of regions is the size of the region map. As shown in Table 6-1, the greater the number of the region the larger the rates of the feedback channel. An appropriate regions number is the greatest number that gives longer bitplane and smaller region map. From Table 6-1 the numbers 4-5-6, seem appropriate, therefore the R-D performance will be investigated for some combinations of these three options (4, 5, 6).

6.2 Region-Based Adaptive DVC Codec: Pixel Domain Test Result

This section presents the results and discussion of the proposed Region-Based adaptive DVC codec in pixel domain presented in section 3.3. Video sequences with different motion activity varying from low to high are used for testing the proposed solution. The performance of the proposed solution is first tested for different regions number for all video sequence. Certain rate points representing the low rate and high rate are chosen to numerically explain the performance gap between the proposed solution and the *state-of-the-art* Frame-Based DVC codec.

6.2.1 RD Performance for Miss America Sequence (Low motion activity)

Figure 6-1 illustrates the R-D performance of the proposed Region-Based Adaptive DVC codec for different region numbers (4, 5, 6). Partitioning the WZ frame into 5 and 6 regions degrades the performance of proposed Region-Based Adaptive DVC codec due to the rate penalty caused by high rate of the feedback channel that is needed to transmit the region map (see Table 6-1). The high number of region also produces shorter bitplane lengths which also degrades the performance of the Slepian-Wolf codec and the overall performance of the proposed Region-Based Adaptive DVC codec.

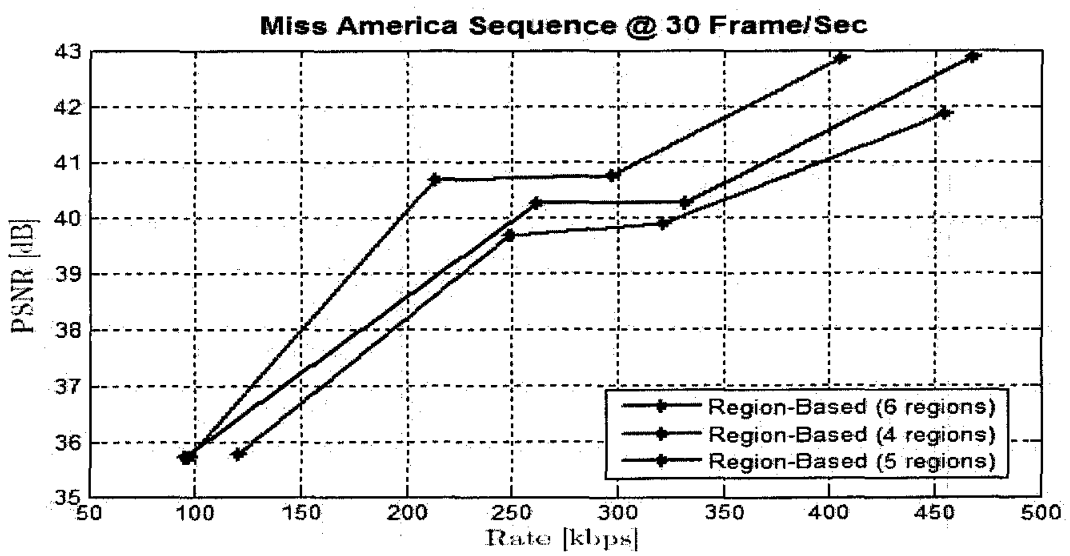

Figure 6-1 R-D performance with different regions number for Miss America sequence

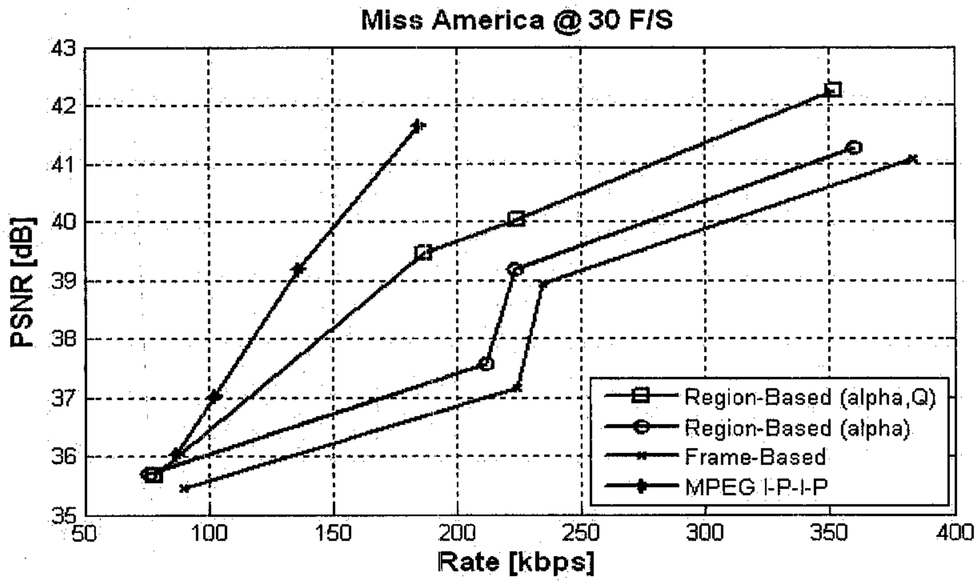

Figure 6-2: RD performance for the Miss America test sequence Pixel domain

Figure 6-2 above shows the RD performances of the proposed Region-Based Adaptive DVC codec with its two implementation cases for the "Miss America" as low motion activity sequence. The first case demonstrates the impact of the accurate dependency channel model (alpha) on the RD performance of the proposed solution, while the second case demonstrates the impact of also including adaptive source coding (Quantizer).

In the first case (alpha), it can be observed that the Region-Based adaptive DCV codec with its accurate dependency channel model feature (alpha) outperforms the Frame-Based DVC in terms of bitrate by saving 17% to 22% for 36 and 41 dB PSNR values. The bitrate saving is achieved by tightening the control over the rate control process and using appropriate dependency channel model parameter (alpha).

In the second case, the performance of the Region-Based adaptive DCV codec is further improved in terms of PSNR (0.4 to 1.3 dB) for the same (89 and 350 kbps) rate values by further including the adaptive source coding feature (Quantizer). This is largely attributed to the process of investing higher rates at the poorly estimated regions and lower rate at the better estimated regions such as the background. Table 6-2 shows the average selected quantizer set corresponding to each rate point. The kink in the two curves is caused by the key frame quantization parameter.

Table 6-2: Average quantizer set for different rate point for Miss America sequence

Fixed Quantizer/ frame (bits)	Region Adaptive Quantizers (bits)			
	Region 1	Region 2	Region 3	Region 4
1	1	1	1	1
2	1	2	2	3
3	2	2	4	4
4	2	4	5	5

It can be observed that there is still a large performance gap when the Region-Based adaptive DVC codec RD performance is compared to the MPEG+ inter-frame coding performance. This superiority of the conventional video coding over the DVC codec is gained by making use of the original reference frame at the encoder so it performs optimum rate control unlike the DVC where the encoder has no access to the reference frame at the decoder (estimate of the WZ frame).

6.2.2 RD Performance for News Sequence (Medium motion activity)

Figure 6-3 illustrates the R-D performance of the proposed Region-Based Adaptive DVC codec for different region numbers (4, 5, 6). As Figure 6-3 shows, partitioning the WZ frame into 5 and 6 regions degrades the performance of the proposed Region-Based Adaptive DVC codec. This is due to the rate penalty caused by high rate of the feedback channel to transmit the region map (see Table 6-1). The high number of regions also produces shorter bitplane lengths which also degrades the performance of the Slepian-Wolf codec and degrades the overall performance of the proposed Region-Based Adaptive DVC codec.

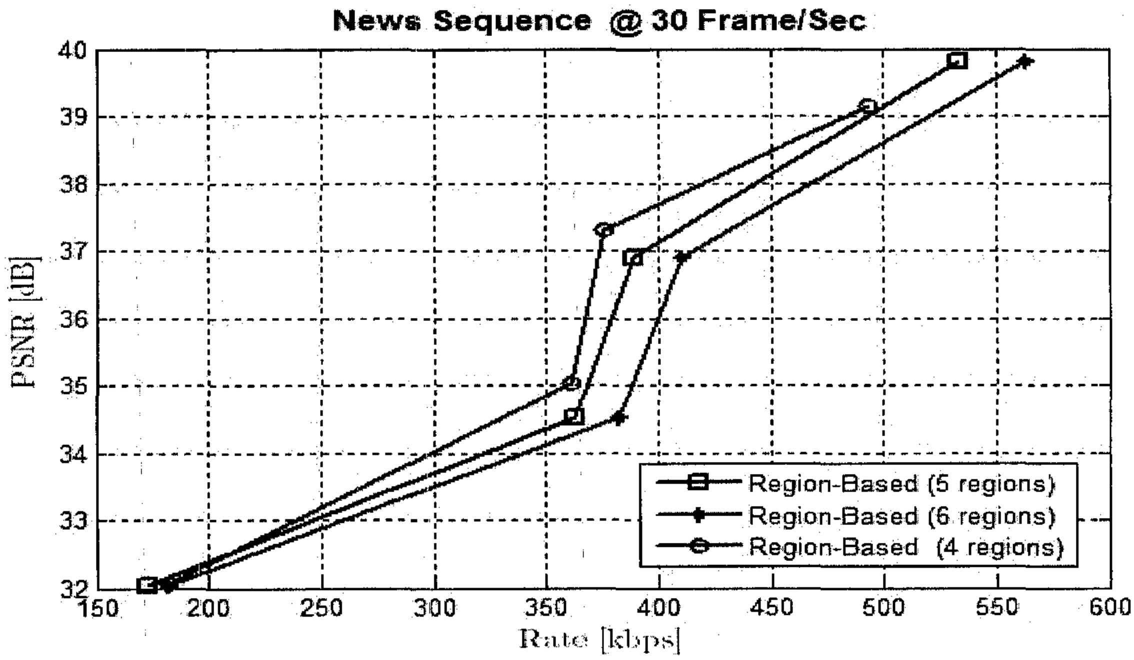

Figure 6-3 R-D performance for proposed DVC codec with different region number for News sequence

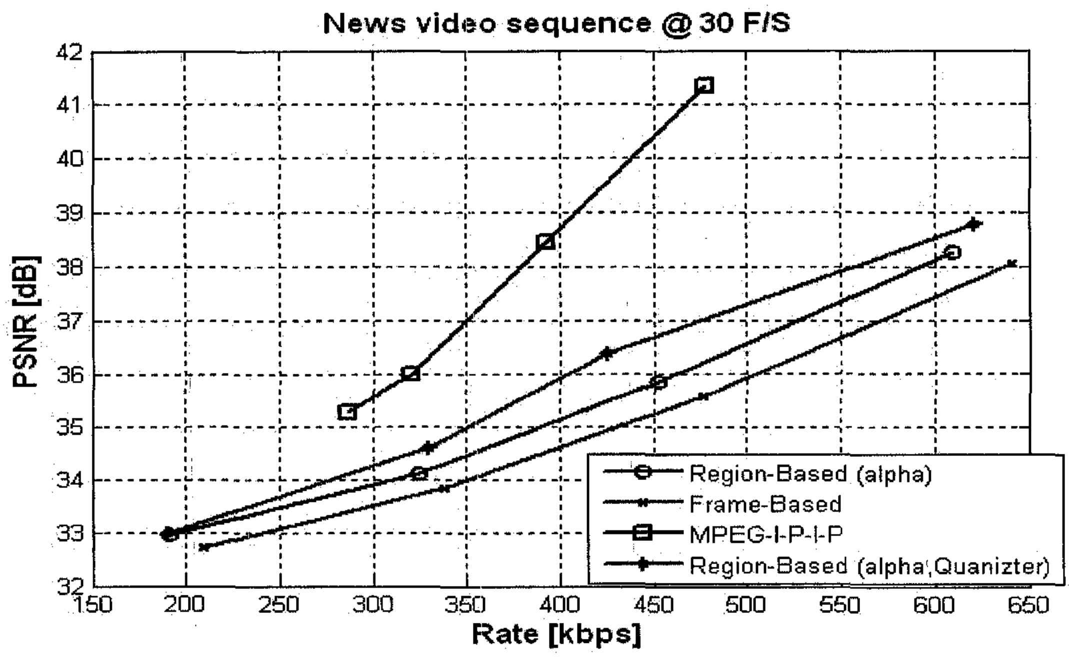

Figure 6-4: RD performance for the News test sequence Pixel domain

Figure 6-4 above, shows the RD performance for MPEG, Frame-Based DVC and the proposed Region-Based Adaptive DVC codec solution for the "News" as medium motion activity sequence. The proposed Region-Based Adaptive DVC codec solution with its adaptive dependency channel model (alpha) and adaptive source coding

(Quantizer) features is implemented in two different cases. In the first case, the proposed Region-Based Adaptive DVC codec solution with its accurate dependency channel model (alpha) feature is shown in Figure 6-4. This feature improves the rate control process and enables the proposed solution to outperform the Frame-Based DVC system by saving 15% to 20% of bitrate for 33 and 38 dB PSNR values. The rate saving can be justified by the better modeling of the dependency channel model (alpha) and tighter control over the rate control process.

In the second case, the proposed Region-Based Adaptive DVC codec with both of its features (Quantizer, alpha) can be seen with PSNR gain from 0.3 to 0.8 dB for 220 and 600 kbps bitrate values. As in the case of RD evaluation test of Miss America, this PSNR improvement is largely attributed to the process of allocating adaptive quantizer depth based on the region characteristics. In other words, allocating higher rates to the poorly estimated regions and lower rate to other regions (background) helps enhance RD performance. Table 6-3 shows the average selected quantizer set corresponding to each rate point.

As before, it can be observed that there is still large performance gap between the proposed solution and the MPEG+ inter-frame coding performance.

Table 6-3: Average quantizer set corresponding to different rate point for News sequence

Fixed Quantizer/ frame (bits)	Region Adaptive Quantizers			
	Region 1	Region 2	Region 3	Region 4
1	1	1	1	1
2	1	2	2	3
3	2	2	4	4
4	3	4	4	5

6.2.3 RD Performance for Mother-Daughter Sequence (Medium motion activity)

Figure 6-5 illustrates the R-D performance of the proposed Region-Based Adaptive DVC codec for different region numbers (4, 5, 6). The R-D of 5 regions cases outperforms the R-D performance of the system when 4 and 6 regions cases. The poor performance of the proposed solution in case of 4 regions is justified by the fact such medium motion activity video sequence requires more regions to cater for its spatial variation in the local characteristics. In the case of 6 regions, the performance degrades due to the rate penalty caused by high rate of the feedback channel to transmit the region map. The high number of region also produces shorter bitplane lengths which also degrades the performance of the Slepian-Wolf codec and the overall performance of the proposed Region-Based Adaptive DVC codec.

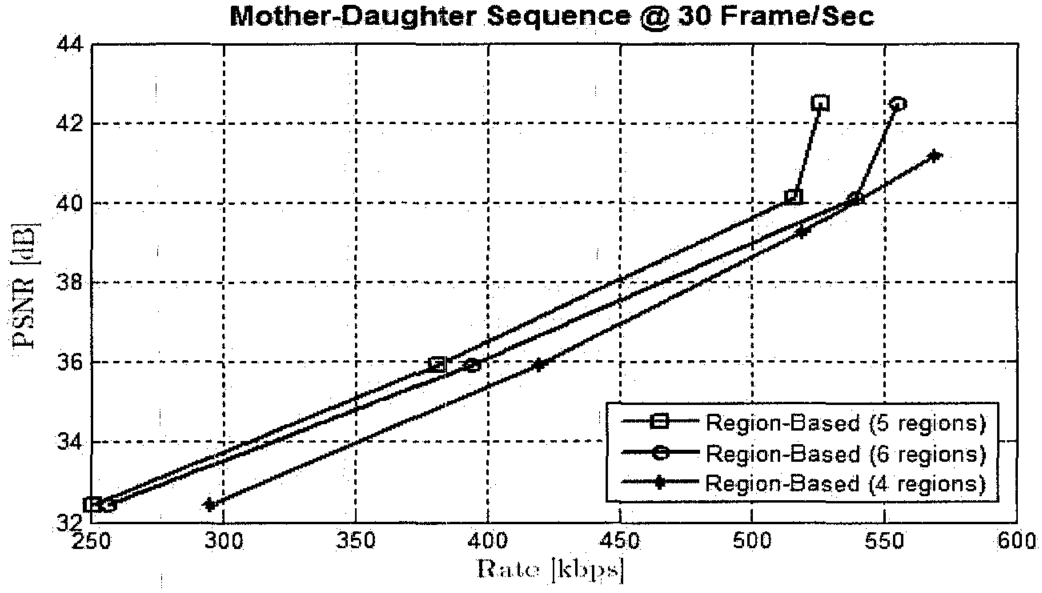

Figure 6-5 R-D performance of different region numbers for the Mother-daughter sequence

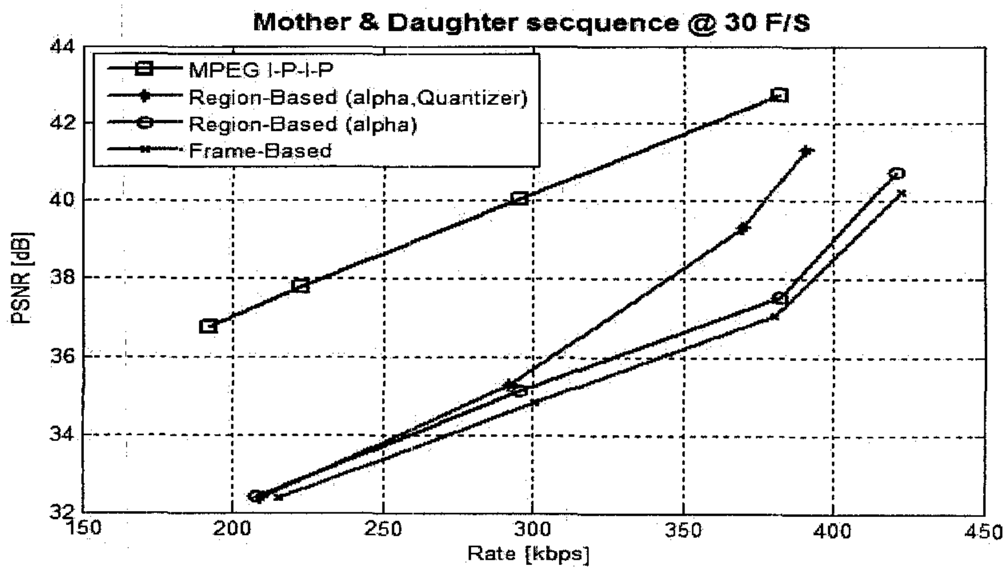

Figure 6-6: RD performance for the Mother-daughter test sequence Pixel domain

Figure 6-6, above shows the RD performance for MPEG, Frame-Based DVC and the proposed Region-Based Adaptive DVC codec solution with its two implementation cases for "Mother-Daughter" as medium motion activity sequence.

In the first case, the proposed Region-Based Adaptive DVC codec is first utilizing only the dependency channel model (alpha) feature. As the figure shows the proposed system outperforms the Frame-Based DVC codec and saves 4% to 8% bitrate for 32.6 and 40 dB PSNR values although extra rate is consumed to communicate the regions map. The bitrate saving advantage is a result of accurate dependency channel model parameter (alpha) and the tight control over the rate control process obtained by the Region-Based video coding.

In the second case, the proposed Region-Based Adaptive DVC codec employs the adaptive source coding (Quantizer) feature. This feature is added to the Region-Based Adaptive DVC codec and the system is tested by adaptively allocating dependency channel model parameter (alpha) and source coding parameter (Quantizer). As shown in Figure 6-6, the Region-Based Adaptive DVC codec performance is further improved and 0.3 to 1.2 dB is gained in PSNR for 231 and 395 kbps bitrate values. As before, this improvement is largely attributed to the process of allocating higher rates to the poorly estimated regions such as the foreground region and occluded regions. Table 6-4 shows the average selected quantizer set corresponding to each rate point. Although these two features enhanced the DVC codec solution but still large performance gap can be observed when its RD performance is compared to the MPEG+ inter-frame coding performance.

Table 6-4 Average quantizer set corresponding to different rate point for Mother-Daughter sequence

Fixed Quantizer/ frame (bits)	Region Adaptive Quantizers				
	Region 1	Region 2	Region 3	Region 4	Region 5
1	1	1	1	1	1
2	1	2	2	3	3
3	2	3	3	4	4
4	3	4	4	5	5

6.2.4 RD Performance for Carphone Sequence (High motion activity)

Figure 6-7 shows the R-D performance of Region-Based Adaptive DVC codec for the Carphone video sequence with different region numbers. Although the higher number of regions for high motion activity video sequence (Carphone) allows finer regions and more appropriate coding parameters allocation, the performance of the Region-Based Adaptive DVC codec when partitioning the WZ frame into 5 regions outperforms the case of 6 regions and 4 regions. This is caused by the high rate of the feedback channel in case of 6 regions as shown in Table 6-1 where transmitting the region map requires more than 0.5 kbps which degrades the R-D performance the proposed Region-Based Adaptive DVC codec. Moreover, the large region number results in shorter size of the bitplane which degrades the performance of Slepian-Wolf codec and contributes to the degradation of the overall R-D performance of the system.

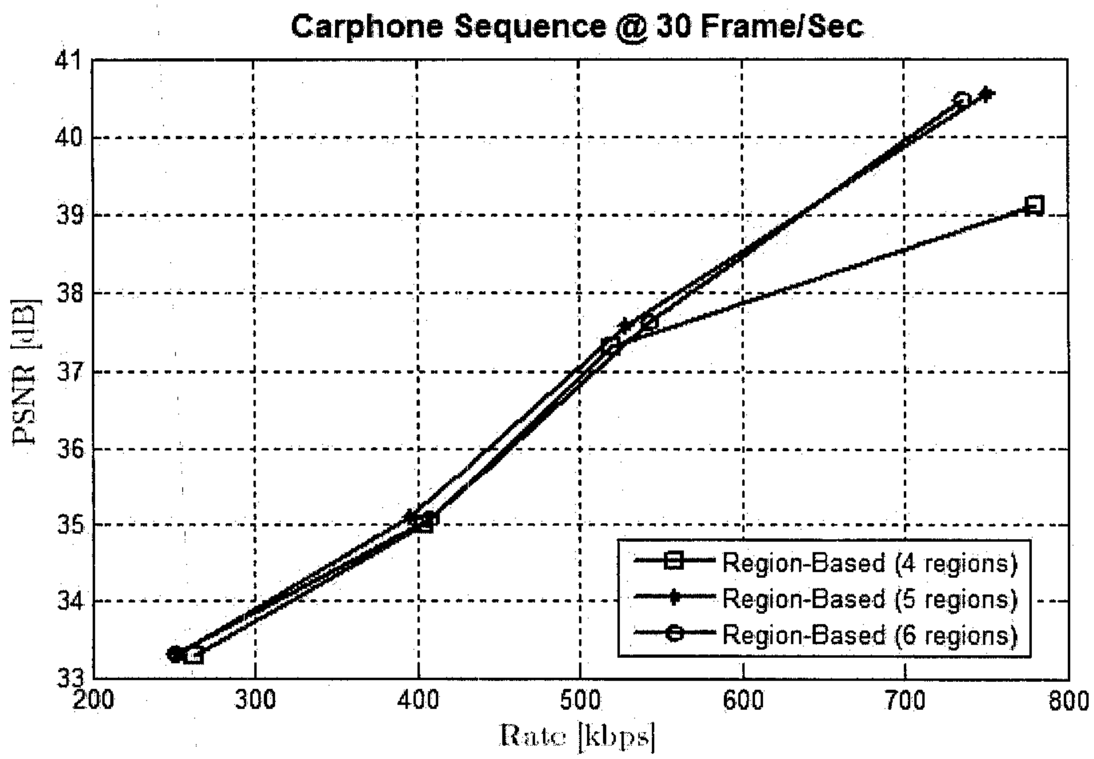

Figure 6-7 R-D performance with different regions numbers for Carphone sequence

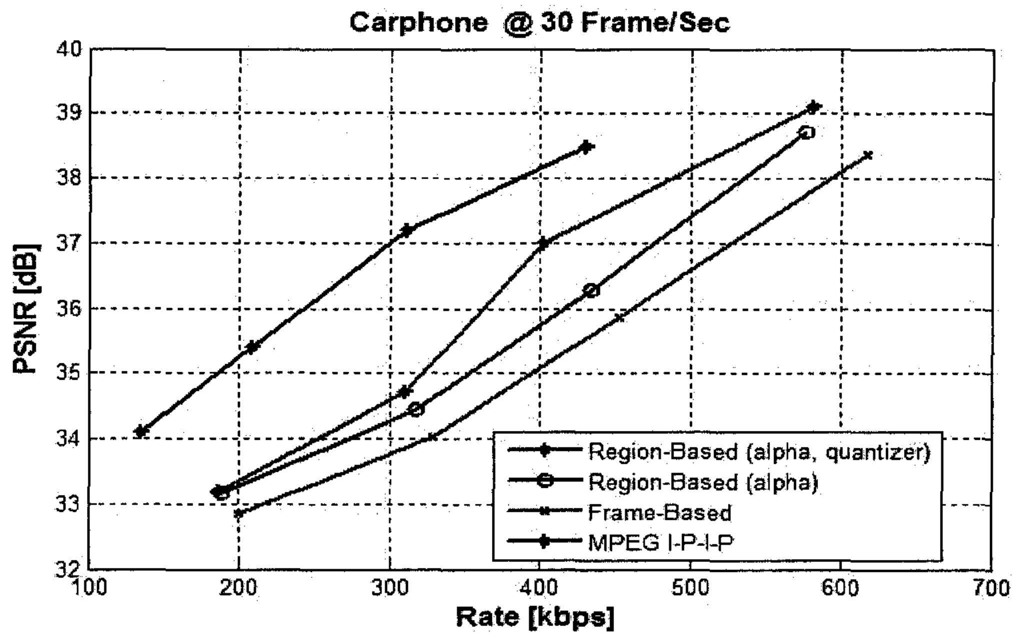

Figure 6-8: RD performance for the Carphone test sequence Pixel domain

Figure 6-8 above shows the RD performances of MPEG, Frame-Based DVC and the proposed Region-Based Adaptive DVC codec solution for "Carphone" as high motion activity sequence in two cases. In the first case, the impact of the accurate dependency channel model (alpha) feature on the RD performance of the proposed solution is evaluated. Similarly, the second case includes the adaptive source coding (Quantizer) to evaluate its impact of on proposed solution RD performance.

In the first case (alpha), shown in Figure 6-8 the proposed Region-Based Adaptive DVC codec outperforms the Frame-Based DVC and enables bitrate saving that varies from 5% to 11 % for 33.4 and 38 dB PSNR values. This improvement is attributed to the process of allocating appropriate dependency channel model parameter (alpha) and to employing the Region-Based coding concept that allows tight control over the rate.

Figure 6-8 also shows that the proposed Region-Based Adaptive DVC codec system performance is further enhanced by employing the second feature adaptive source coding too (Quantizer). The proposed solution achieves PSNR gain varying from 0.3 to 0.9 dB for 200 and 585 kbps bitrate values respectively. This gain is largely attributed to the process of allocating quantizer depth based on the local characteristics of the video sequence. In other word, the WZ frame is quantized region-by-region and only the region which is poorly estimated requires finer quantization steps. Table 6-5 shows the average selected quantizer set corresponding to each rate point.

Although these features enhance the RD performance of the DVC codec when compared to the Frame-Based codec, the figure also shows that the MPEG+ codec still outperforms the proposed DVC codec. This outperformance is gained because MPEG+ encoder has access to the reference frame.

Table 6-5 Average Quantizer Set Corresponding To Different Rate Point for Carphone Sequence

Fixed Quantizer/ frame (bits)	**Region Adaptive Quantizer**				
	Region 1	Region 2	Region 3	Region 4	Region 5
1	1	1	1	1	1
2	1	2	2	3	3
3	2	3	3	4	4
4	3	4	4	5	5

6.2.5 Remarks on Region-Based Adaptive DVC Codec in Pixel Domain

The results presented in section 6.2, show that the RD performances of Region Based Adaptive DVC codecs in Pixel domain are considerably higher in terms of RD performance than the state-of-the-art frame-based DVC codec for all bitrates and all the test sequences. Although the performance of the proposed Region-Based Adaptive DVC codec varies from one sequence to other, it can be observed from the results that it performs better in terms of RD performance for the sequence with low/medium motion activity, since the number of regions approximately matches the number of moving objects with those sequence (Miss America, News). The rate of the feedback channel degrades the overall R-D performance of the proposed codec (Mother& Daughter, Carphone). Therefore, the proposed scheme deliberately does not use more regions as it causes more bits from the region map to be transmitted back, resulting in rate penalty. It is also desirable that the number of bits from any given bit plane is large enough so the Turbo codes used for generating the parity bits work efficiently. This requirement restricts the number of regions to be no more than 4 or 5. However, comparing RD performance of the Region Based Adaptive DVC codec to the MPEG+ interframe coding with I-P-I-P structure, it can be seen that there is still a performance gap.

Even though both MPEG+ and proposed DVC technique try to make use of the fact that each frame, from coding perspective, comprises of different coding regions, the outperformance by the MPEG is because the reference frame is available to its encoder while the proposed DVC codec only makes do with an estimate of the reference frame and that too only at the decoder. The lack of restriction on encoding complexity in MPEG+ allows the encoder to also build the decoder at the encoder and to predict the reconstruction error so as to afford adaptive quantization and tighter rate control. This cannot be done in DVC solutions, nor can it be done in the proposed technique, hence the performance gap. This performance gap is more for video sequences characterized by higher amount of motion since MPEG+ schemes uses large number of regions while the proposed solutions is constrained to use small numbers of the region.

6.3 Test Results of REGION-BASED DVC Codec in Transform Domain

This section presents the performance results of the proposed Region-Based adaptive DVC codec in transform domain presented in section 3.4. The same video sequences used in testing the proposed solution in pixel domain are used for testing the proposed solution in transform domain. The performance of the proposed solution is tested using appropriate number of regions identified in the testing of the proposed solution in pixel domain (4 for low motion and 5 for high motion activity). Certain rate points representing the low rate and high rate are chosen to numerically present the performance of the proposed solution vis. a vis. the state-of-the-art Frame-BASED DVC codec in the transform domain. In this discussion, the R-D performance improvement over frame-based DVC is presented in two successive stages - first, the impact of using only the accurate Region-Based dependency channel model and secondly, the impact of including adaptive source coding (Quantizer) in the proposed solution. The proposed solution with accurate dependency channel model (alpha) outperforms the Frame-Based DVC codec by saving 27%-44% and 26-37% bitrate for same PSNR values for "Miss America" and "News" as shown in Figure 6-10 and

Figure 6-11 respectively. The performance of the proposed solution outperforms the Frame-Based DVC Codec for medium high and high motion video sequence by 28-42% "Mother-daughter" and 33-37% "Carphone" sequences as shown in Figure 6-12 and figure 6-9 respectively. This improvement can be justified by the use of more accurate dependency channel modeling which enhances the performance of Slepian-Wolf codec (Turbo code). In the Region-Based coding the WZ frame is decoded region-by-region and the successfully estimated region requires less bitrate so the rate control is tight.

The performance of the proposed solution when the adaptive source coding in the form of quantizer depth selection is also included shows that the proposed solution achieves PSNR gain from 0.8 up to 2.2 dB and 1 to 2.1 dB for same lower and higher bitrates values for "Miss America" and "News" as shown in Figure 6-10 and Figure 6-11 respectively. Including the feature of allocating appropriate source coding results in 1.5-2.3 dB PSNR gains for the same bitrate values for "Mother-Daughter" and "Carphone" sequences as shown in Figure 6-12 and figure 6-14 respectively. The PSNR gain is introduced to the performance of the proposed solution by adaptively allocating the quantization matrix that considers the local characteristics of the video sequence. In other words, the adaptive quantization selection allocates finer quantizers to poorly estimated region such as the occluded region.

From the results depicted in figures 6-11 to 6-15, it can be observed that there is still a compression gap when compared to the MPEG+ interframe coding with an I-P-I-P structure. This performance gap is attributed to the availability of the reference frame at the encoder in MPEG+ that enables tighter control of bitrate as explained in section 6.2.5.

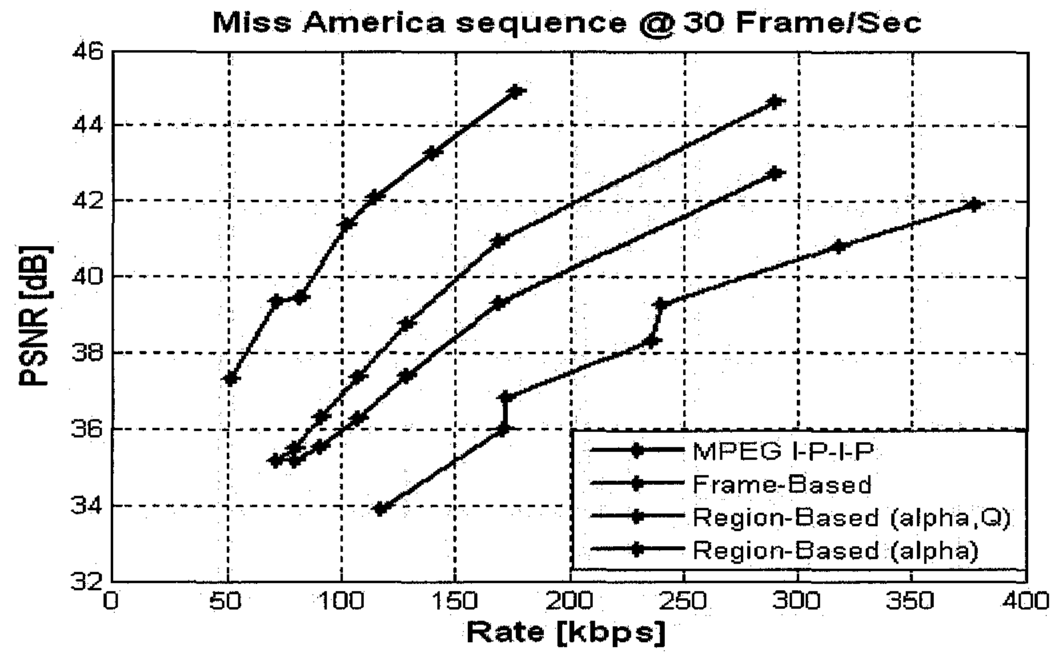

Figure 6-10 RD performance for the Miss America test sequence Transform domain

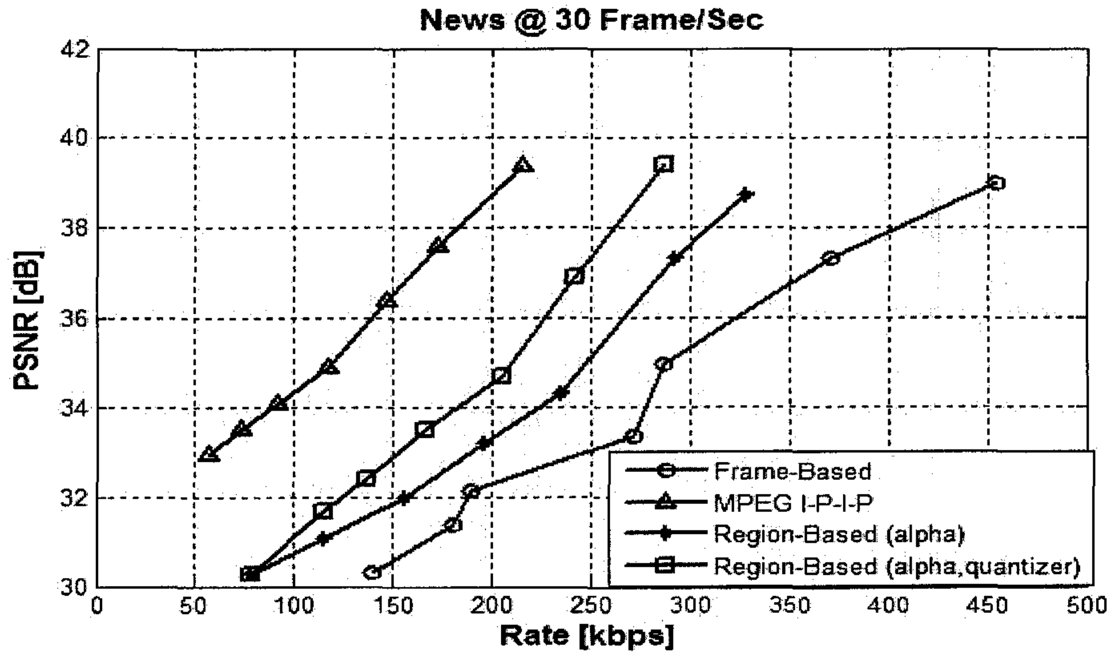

Figure 6-11: RD performance for the News test sequence Transform domain

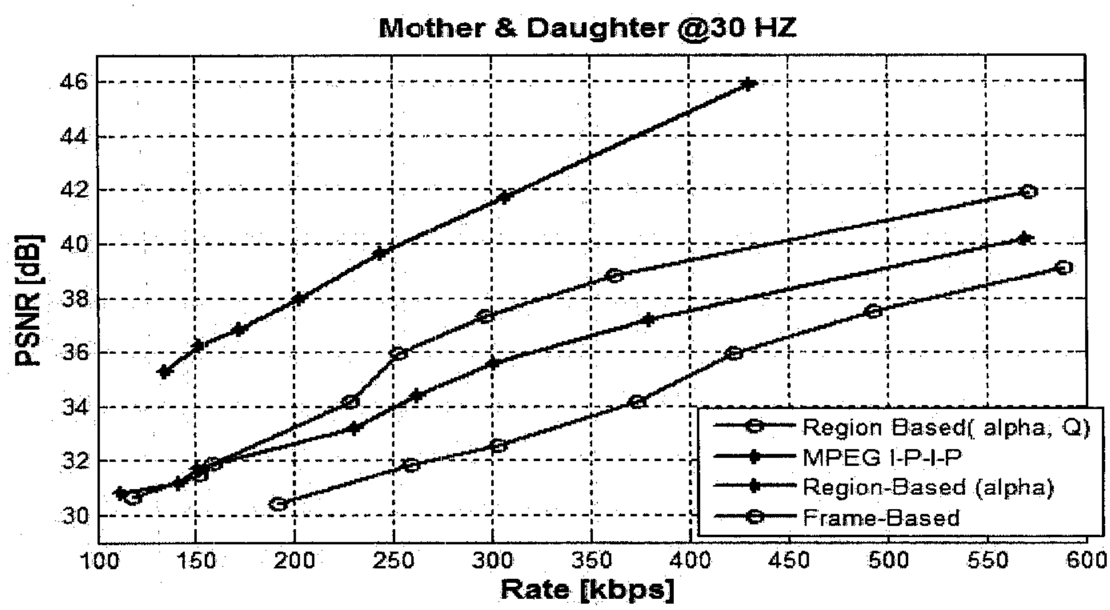

Figure 6-12: RD performance for the Mother-Daughter test sequence Transform domain

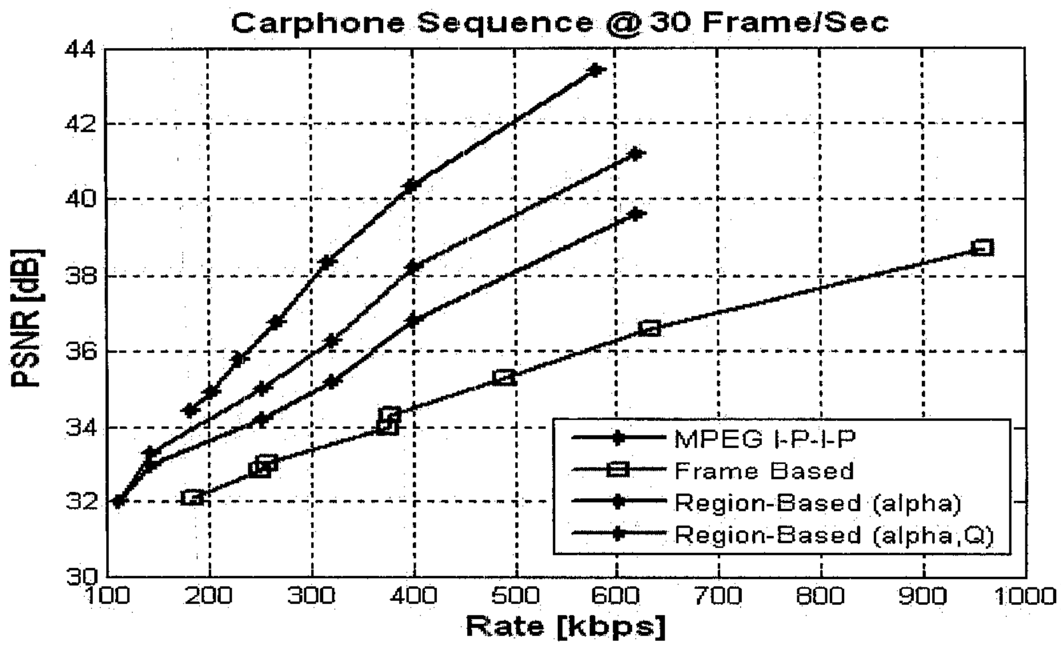

Figure 6-13: RD performance for the Carphone test sequence Transform domain

6.3.1 Remarks on Region-Based Adaptive DVC codec in Transform Domain

The results presented in section 6.3 show that the RD performance of Region Based Adaptive DVC codec in transform domain is considerably above the state-of-the-art frame-based DVC codec transform domain for all bitrates and all the test sequences. The results show that the Region Based Adaptive DVC codec in transform domain RD performance outperforms the Region Based Adaptive DVC codec in pixel domain. This outperformance is a result of using DCT in every frame to exploit spatial redundancy. The performance in Region-Based Adaptive DVC codec in transform domain for video sequences with low/medium motion activity is better than its performance in the high motion activity sequences. This is due to allocating appropriate number of regions (4/5) for such sequences. The proposed scheme deliberately does not use more regions as it causes more bits from the region map to be transmitted back, resulting in rate penalty. It is also desirable that the number of bits from any given bit plane is large enough so the Turbo codes used for generating the parity bits work efficiently.

However, comparing RD performance of the Region Based Adaptive DVC codec to the MPEG+ interframe coding with I-P-I-P structure, it can be seen that there is still a compression gap. This gap is more for video sequences characterized by high amount of motion. This gap is the result of the MPEG+ encoder's ability to perform better rate control since the reference frame is available at the encoder as highlighted in section 6.2.5.

6.4 Region-Based Adaptive DVC codec without feedback in Pixel Domain: Simulation Results

This section presents the results and discussion of the proposed feedback-free Region-Based adaptive DVC codec in pixel domain presented in CHAPTER 4. Once again, same video sequences used in testing the proposed solution in pixel domain are used for testing the proposed solution. The performance of the proposed solution is tested for appropriate number of regions identified in the testing of the proposed solution with feedback in pixel domain (section 6.2). The proposed DVC codec makes use of two different source coding schemes – one with fixed number of quantizer bits for all regions and two, with average quantizer sets determined in section 6.2. Certain rate points representing low and high rates are chosen to numerically present the performance gap between the proposed feedback-free solution and the Region-BASED DVC codec with feedback based solution.

The RD performance for Region-Based Adaptive DVC codec with feedback and MPEG+ codec are compared to that of the Region-Based Adaptive DVC codec without feedback Pixel Domain solution for the "Miss America" and "News" and "Mother-daughter" sequences. The process of adaptive quantizer depth selection, as it was done in the Region-Based with Feedback DVC codec, is completely removed from the Region-Based without Feedback DVC codec to avoid adding extra complexity to the encoder. The appropriate quantizer's depths are instead obtained from the Table 6-2 to 6.4, where the average quantizer set corresponding to different rate points is obtained adaptively for each video sequence. Appropriate quantizer depth allocation leads to R-D performance improvement as shown later.

The performance of the feedback based solution outperforms the proposed Region-Based Adaptive DVC codec without feedback solution in pixel domain by 0.2-1 dB for "Miss America" and by 1.4 dB for "News" and by 0.2-0.7dB for "Mother-daughter" as shown in Figure 6-15 and Figure 6-15 and Figure 6-15 respectively. The poorer performance of the proposed solution when compared with the feedback based solution is a result of under/over estimation of appropriate rate class. Underestimating of appropriate rate class degrades the overall performance and results in decoding error greater than (10^{-3}), in these cases, the corresponding bitplane at the decoder is used to replace the erroneous decoded bitplane. Because the classifier favors the higher rate classes to avoid the underestimation, the performance of the proposed solution suffers from rate penalty, specifically when finer quantizer is used, since the behavior of the feedback channel for the LSBs is as accurately modeled as needed and is difficult to predict. Figure 6-15 shows that there is rate-saving in the second case caused by using small quantizer depth (first region, background, occupies 70% of the frame) and therefore less bitplane to encode.

In addition, as before, it can be observed that there is still a large gap in comparison with the MPEG+ inter-frame coding performance, the explanation for which is same as in section 6.2.5.

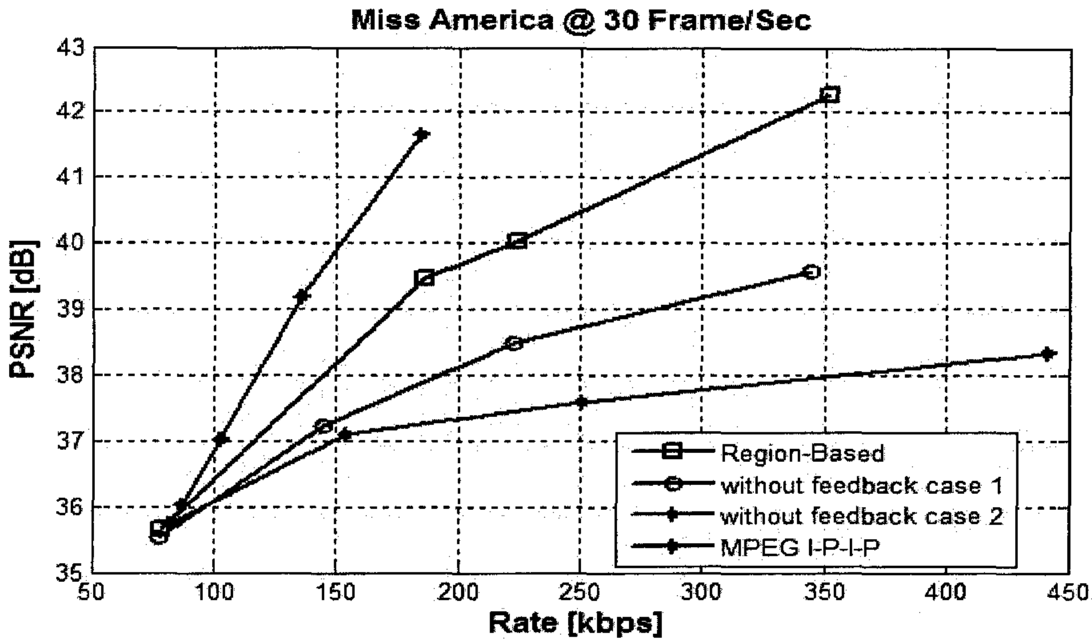

Figure 6-14: RD performance for the Miss America test sequence Pixel domain

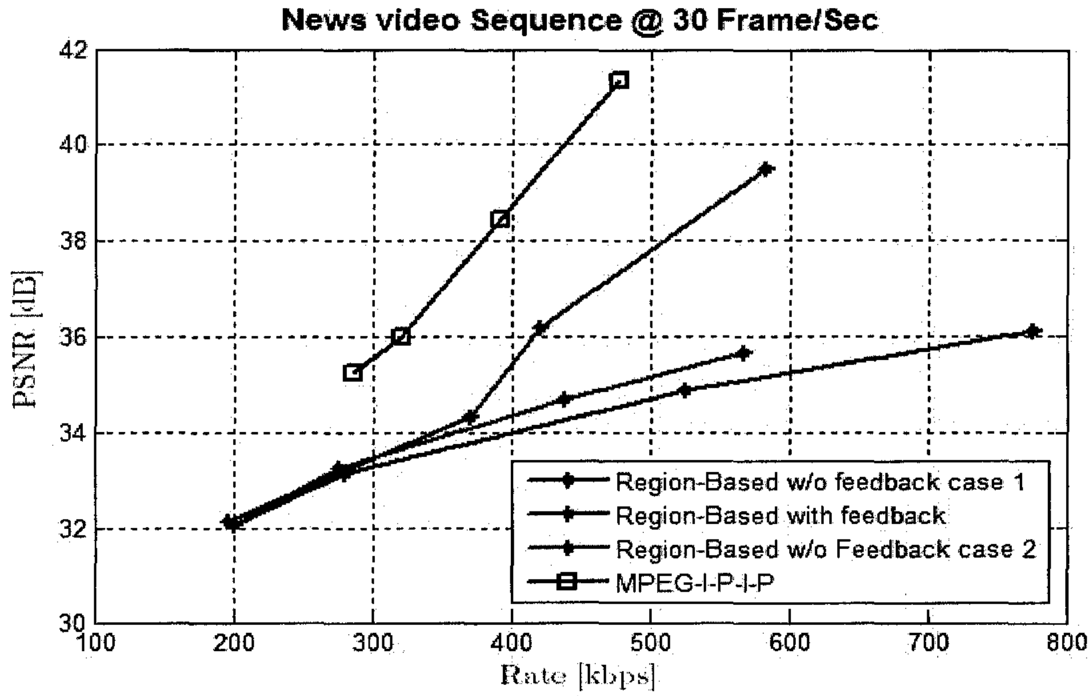

Figure 6-15: RD performance for the News test sequence Pixel domain

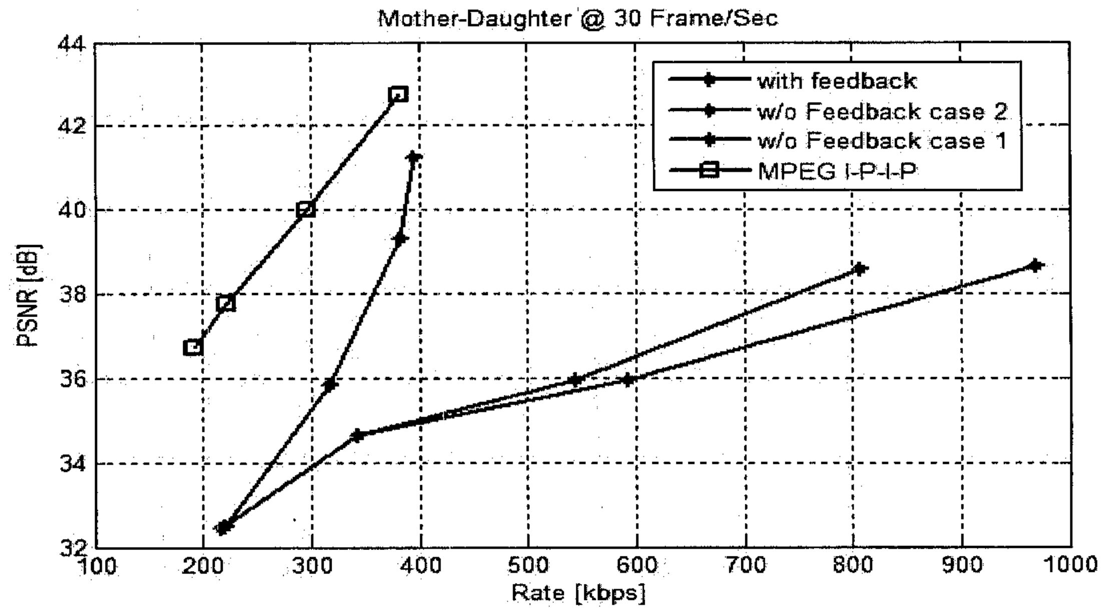

Figure 6-16: RD performance for the Mother-Daughter test sequence Pixel domain

6.4.1 Remarks on Region-Based Adaptive DVC codec without feedback in pixel domain

The results presented in section 6.4 show that the RD performance of the pixel domain Region Based Adaptive DVC codec solution with feedback channel outperforms that of pixel domain Region Based Adaptive DVC codec without feedback channel. The proposed Region Based Adaptive DVC codec without feedback suffers from rate penalty since the rate estimation tends to overestimate the rate to avoid under estimation. Although the overall design of the rate estimation module favors the overestimation but the system suffers from underestimated cases whereby the overall R-D performance of the systems is degraded. When source coding is included in the solution, the R-D performance of the system improves by performing rate allocation based on some offline experiment that helps allocate the appropriate quantizer depth to suit the region characteristics.

However, comparing RD performance of the Region Based Adaptive DVC codec to the MPEG+ interframe coding with I-P-I-P structure, it can be seen that there is still a compression gap. This gap is more for video sequences characterized by high amount of motion. This gap is a result of the encoder ability to perform better rate control since the reference frame is available at the encoder as explained in section 6.2.5.

6.5 General Remarks and Summary

This chapter presents the results of performance evaluation of the proposed Region-based adaptive DVC codec solutions both in pixel and transform domains when a feedback channel is assumed present. The performance of the proposed DVC solution is seen to be better than the state-of-the-art Frame based DVC solution. This outperformance is uniform across all types of video sequences – low or high motion content. This is because the proposed solution divides the incoming video sequences - sequences that are known to consist of more than one coding units, each with its own coding parameters, into appropriate number of regions and allocates appropriate coding parameters. Although there is improvement in the R-D performance, the conventional video coding MPEG+ still outperforms the DVC codec. This is due to the fact that, in MPEG+ schemes, a reference frame is available at the encoder that helps the encoder allocate appropriate coding parameters and obtain better R-D performance, while, in the DVC codec, only an estimate of the reference frame is available, that too, at the decoder, and obtained by frame interpolation techniques. For DVC encoder to make use of the availability of reference frame for appropriate allocation of coding parameters, these computations will have to be done at the decoder and passed on to the encoder with some rate-penalty and channel model inaccuracy.

The performance of the proposed DVC codec can be further enhanced if the frames are free to be partitioned into as many regions as warranted without increasing the rate of the feedback channel. In addition, the performance can also improve if Slepian-Wolf coding continues to support the resulting shorter bitpalnes. However, in the present form, the proposed solution is limited in scope as it cannot partition the frames into more than 5 regions without incurring unaffordable rate penalty.

Furthermore, this chapter has also studied the performance of a Region-Based adaptive DVC codec in pixel domain where the feedback channel is not present. The proposed solution represents good solution in such uni-directional application scenarios. The performance of the proposed feedback-free solution is poorer when compared to the feedback based solution. The rate control based on feedback helps Region-Based adaptive DVC codec with feedback solution to avoid both the over/under estimation of the rate classes' problems. However, that is not possible because the proposed feedback-free solution depends heavily on the training data, resulting in slight degradation of the R-D performance of proposed feedback-free solution. The performance of Region-Based adaptive DVC codec without feedback can be enhanced by refining the encoder version of the side information without adding too much complexity. Another factor that can help decrease the over/under estimation of the rate classes is employing smarter machine learning algorithm and identifying more features to obtain better insight about the behavior of the feedback channel.

CHAPTER 7
CONCLUSION & FUTURE WORK

In this chapter, all research work presented in this thesis will be summarized in section 7.1 and concluded along with its overall contributions in section 7.2. After that, all possible modifications which could improve the performance of the proposed DVC solution will be presented in section 7.3.

7.1 Proposed Solution Summary

Recently new emerging applications such as wireless video sensor networks and multimedia sensor networks, wireless PC cameras and mobile camera phones, represent a real challenge for the conventional video coding architecture. These applications have very different requirements than those of traditional video delivery systems. For some applications, when there is a high number of encoders and only one decoder, e.g. surveillance, low cost encoder devices are necessary. In order to fulfill these requirements, it is essential to have a low-power and low-complexity encoding device, possibly at the expense of a higher complexity decoder. As a result, a lower complexity encoder could be achieved by moving some of the encoder tasks to the decoder part, specially the most complex motion estimation process. This useful hint is induced from a new coding scheme referred to as distributed source coding (DSC) algorithm proposed by Slepian and Wolf for distributed lossless coding and by Wyner and Ziv for lossy coding, discussed in chapter 2. New video coding paradigm based on the DSC known as distributed video coding (DVC). The basic idea of this DVC architecture is that the decoder, based on some previously and conventionally transmitted frames, the so-called key frames, creates the so-called side information (SI) which works as estimates for the other frames to be encoded, the so-called WZ frames.

The WZ frames are then encoded using a channel coding approach. A turbo code or Low-Density Parity-Check (LDPC) codes can be used to correct the 'estimation' errors in the corresponding decoder estimated frames. One can use the previous frame as current frame estimate and request only as few parity bits as are necessary to correct the estimation errors, thus causing huge compression gains. In the case of low temporal correlation between video frames a better estimation to the current frame can be estimated from the previously decoded video frames. By exploiting temporal correlation in video sequences, motion estimation (ME) algorithms are used to predict the current video frame in the sequence. This way, the complex task of ME is shifted to the decoder. Although more sophisticated ME algorithms can be used, the practical frame prediction is fundamentally faulty due to events like occlusions. Therefore, the frame prediction is expected to have spatially varying degrees of success across the predicted frame. Many of the current DVC schemes focus on how to improve the frame estimation process, to avoid the consequences of the frame estimation limitations. An efficient DVC codec must consider this spatial variation along the predicted frame.

This work proposed partitioning the considered frame into many coding units (region) where each unit is encoded differently. This partitioning is, however, done at the decoder while generating the side-information and the region map is sent over to the encoder at very little rate penalty. The partitioning allows allocation of appropriate DVC coding parameters (virtual channel, rate, and quantizer) to each region. The adaptive allocation to the relevant DVC coding parameters helps the proposed solution to collectively address all the Frame-Based DVC codec issues. The resulting regions map is represented by using quadtree algorithm and communicated to the encoder via the feedback channel. The rate control in DVC is performed by channel coding techniques (turbo codes, LDPC, etc.). In this work, a turbo code has been used and Region-Based adaptive DVC codec is designed both in transform domain and in pixel domain. The transform domain video coding (TD) has distinct superior performance as compared to the normal Pixel Domain (PD), since it exploits the spatial redundancy during the encoding. The performance evaluations show that the proposed system is superior to the existing distributed video coding solutions. Although the proposed system requires extra bits representing the "regions map" to be transmitted back to the encoder, still the rate gain is noticeable and it outperforms the state-of-the-art frame based DVC by 0.6-1.9 dB.

The feedback channel (FC) has the role to adapt the bit rate to the changing statistics between the side information and the frame to be encoded. In the unidirectional scenario, the encoder must perform the rate control. To correctly estimate the rate, the encoder must calculate typical side information. However, the rate cannot be exactly calculated at the encoder, instead it can only be estimated. In this work, we also propose a feedback-free region-based adaptive DVC solution in pixel domain based on machine learning approach to estimate the side information. Although the performance evaluations show rate-penalty, it is still acceptable considering the simplicity of the proposed algorithm.

7.2 Contributions

The main contribution of this work is the design of a DVC codec that considers and makes use of the local characteristics of the video sequence. The proposed DVC solution adheres to the low-encoding-complexity constraint by performing most of the tasks at the decoder. To avoid adding extra complexity to the encoder, the solution reutilizes the existing state-of-the-art components, such as the side information generation and the feedback channel in case of bi-directional application scenario. The region map interpretation at the encoder to partition the WZ frame adds very negligible complexity. In the case of uni-directional application scenario, the complexity constraints are adhered to by using simple machine learning algorithm to estimate the rate at the encoder based on a-priory learnt features during a set-up phase. The proposed Region-Based Adaptive DVC codec solutions in the two application scenarios have specifically made following contributions to the DVC codec research field:

7.2.1 Region-Based Adaptive DVC Codec with Feedback Contributions

I. The Region Based overcomes the problem of inaccurate dependency channel model; by incorporating a location-specific noise model for each region. The dependency channel is estimated per region to avoid the stationary model for the entire frame; which is equivalent to allocating different channel, for each region.

II. The Region based enhanced the rate control process; even though the encoder "blindly" sends the parity bits; the locations of decoding errors are limited into one region rather than a whole bitplane. Thus, the parity bits sent are more accurate. It also allocates different levels of protection is for each region, for instance the high motion area causing failure of decoding only belongs to one region. The decoder requests more parity bits only when this region is sent. Before that, for other regions like a region only containing the background information, the decoding will be successful and no requests for more parity bits will happen. Experimental results show that this Region-Based approach delivers better estimates of the adequate encoding rate than the frame-based approach

III. The proposed region based adaptively allocates a rate based on the local video characteristics. This allows applying appropriate source coding for each region targeting a certain operational RD point. The reconstruction distortion is further reduced since the quantizer is selected based on the initial estimation distortion of the region. As a result, a finer quantizer is allocated to the poorly estimated region. Experimental results show that this adaptive quantization selection region-based approach enhances the overall PSNR when compared with a fixed quantization region-based approach.

7.2.2 Uni-directional Scenario - "Region-Based Adaptive DVC Codec without Feedback" Contribution

The second major contribution is in the design of Region-Based adaptive DVC codec without feedback. The proposed DVC system retains the design philosophy of region based DVC codec of bi-directional scenario. The design is made possible with following contributions:

I. Using a *simple averaging technique*, the encoder partitions the WZ frame into different regions.

II. Each region is recognized as one of the region-type. An *offline study* is separately conducted that classifies the feedback channel of Bi-directional Region Based DVC design into certain rate classes and associates these rate classes to a specific region-type.

III. The encoder uses an appropriate source coding scheme for each region and sends the region map and the quantizer sets unilaterally to the decoder which can then successfully reconstruct the frame.

7.3 Future Work

7.3.1 Better Estimate of Side Information

The quality of side information has great influence on the overall performance of the system. Therefore improving and refining the side information can further enhance the performance of the proposed Region-Based Adaptive DVC codec.

7.3.2 Mixed Mode DVC

The spatial variation of the reliability of the side information predication can further be utilized by partitioning the side information into regions; where some regions are skipped such as the background parts. A better performance can be achieved by intra-coding some regions where the estimation reliability is very small.

7.3.3 Larger GOP

This work has only considered the case when GOP=2; studying the performance of the proposed system when larger a GOP is used. A special frame partitioning procedure need to be devised for large GOP, considering that the previous regions maps in composing the current frame regions map can be performed at the encoder by devising simple method to maintain the low-complexity at the encoder. Also study needs to be conducted to study the possibility of using the same regions map for different frames and associate the GOP size with regions map.

7.3.4 Simple Side Information Algorithm

The rate control work presented in this thesis uses a simple algorithm to estimate the side information at the encoder. The quality of the encoder version of the side information needs to be improved to be a reliable in the rate estimation process. To improve the quality of this version and to maintain the encoder simplicity at the same time, a simple side information algorithm can be useful.

7.3.5 Multiview Region-Based DVC Codec

Multiview DVC incorporates all single video DVC components and differs from the single view case only at side information generation. More precisely, the side information is constructed not only using the frames within the same camera but using frames from the other cameras as well. In a Multiview DVC scenario with no inter-camera exchange disparity estimation must be performed by the decoder. Capturing inter-view dependency turns out to be more difficult than for temporal dependencies, as, in general, multi-view images contain disparities much larger than displacements between successive frames. The regions of one single scene can have their highest correlation with regions from different scene based on the local region characteristics. Such scenario can benefit from the Region-Based coding by allocating appropriate DVC coding parameters for each region based on its local characteristics. Hence, the Region-Based Adaptive DVC codec for Multiview video coding scenario seems promising and it can be extended.

In conclusion, this thesis has demonstrated that a Region-Based DVC codec is feasible and it can offer significant improved performance.

REFERENCES

[1] A. Aaron, S. Rane, and B. Girod, "Wyner-Ziv video coding with hash-based motion compensation at the receiver," in *IEEE International Conference on Image Processing* Singapore: IEEE, 2004.

[2] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video "Invited paper"," in *Proc. Asilomar Conference on Signals and Systems* Pacific Grove, CA: IEEE, 2002.

[3] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," in *Proc. IEEE, Special issue on advances in video coding and delivery*, pp. 1-83, 2005

[4] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions. Information.Theory,* vol. 19, pp. 471-482, july 1973.

[5] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory,* vol. 22, pp. 1-10, 1976.

[6] R. Puri and K. Ranichandran, "PRISM: A "Reversed" Multimedia Coding Paradigm," in *Image Processing International Conference*, Sardinia, Italy, , pp. 617-20, 2003

[7] A. Aaron, E. Setton, and B. Girod, "Towards practical Wyner-Ziv coding of video," in *IEEE International Conference on Image Processing*, Barcelona, Spain, 2003.

[8] A. Aaron and B. Girod, "Wyner-Ziv video coding with low-encoder complexity "Invited Paper"," in *Proc. Picture Coding Symposium* San Francisco, CA, 2004.

[9] J. E. Fowler, "An Implementation of PRISM Using QccPack," Mississippi State University 2005.

[10] R. Puri and K. Ramchandran, "PRISM: an uplink-friendly multimedia coding paradigm," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.*

(ICASSP '03). 2003 IEEE International Conference on, , pp. IV-856-9 vol.4, 2003

[11] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. IEEE Data Compression Conference*, Snowbird, UT, 2002.

[12] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inform. Theory,* vol. 49, pp. 626-643, March 2003.

[13] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," in *Visual Communications and Image Processing* San Jose, CA,: IEEE, 2004.

[14] J. Bajcsy and P. Mitran, "Coding for the Slepian-Wolf problem with turbo codes," in *IEEE Global Telecommunications Conference*: IEEE, pp. 1400-1404 , 2001

[15] R. P. Westerlaken, S. Borchert, R. K. Gunnewiek, and R. I. L. Lagendijk, "Finding a near optimal dependency channel model for a ldpc-based wyner-ziv video compression scheme," in *12th annual conference of the Advanced School for Computing and Imaging,* Lommel, Belgium, 2006.

[16] P. F. A. Meyer, R. P. Westerlaken, R. K. Gunnewiek, and a. R. L. Lagendijk, "Distributed source coding of video with non-stationary side-information," in *Proc. Visual Communications and Image Processing* 2005.

[17] J. Haghhighat, "Data Compression using Error Correcting Codes," in *Electrical and Computer Engineering*. vol. PhD Montreal: Concordia University, p. 100, 2007

[18] J. Venkataraman, "Analysis Of Slepian Wolf Coding," in *Electrical Engineering*. vol. Master Notredame: University of Notredame, p. 111, 2004.

[19] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed Source Coding for Sensor Networks," *IEEE Signal Processing Magazine,* vol. 2, pp. 421-440, 2004.

[20] A. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inform. Theory,* vol. 20, pp. 2-10, 1974.

[21] C. Berrou and A. Glavieux, "Near Optimum Error Correcting Coding and Decoding:Turbo-codes," *IEEE Transactions. Communication* vol. 44, pp. 1261-1271., 1996.

[22] A. Shokrollahi, "LDPC Codes: An Introduction," Digital Fountain, Inc.

[23] D. Schonberg, S. Pradhan, and K. Ramchandran, "Distributed code constructions for the entire Slepian-Wolf rate region for arbitrarily correlated sources," in *IEEE* Snowbird,UT: IEEE, 2004.

[24] V. Stankovic, A. D. Liveris, Zixiang Xiong, and Costass, "Design of Slepian-Wolf Codes by Channel Code partitioning," in *IEEE Proceedings of Data compression Conference*. vol. 04 WASHINGTON: IEEE pp. 302-311, 2004.

[25] Coleman, T. P. Lee, A.H. Medard, and M. M. Effros, "On some new approaches to practical Slepian-Wolf compression inspired by channel coding " in *IEEE Proceedings of Data compression Conference*, WASHINGTON,USA, pp. 282- 291, 2004

[26] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes,," in *IEEE Communications Letters*. vol. 5: IEEE, pp. 417-419,2001.

[27] Q. Xu and Z. Xiong, "Layered Wyner-Ziv Video Coding," *Image Processing, IEEE Transactions on* vol. 15, pp. 3791-3801, 2006.

[28] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. Inform. Theory,* vol. 50, p. 18, 2004.

[29] L.Bahl, J.Cocke, F.Jelinek, and J.Raviv, "Optimal Decoding of Linear Codes for minimizing symbol error rate," *IEEE Transactions on Information Theory,* vol. IT-20, pp. 284-287, March 1974. 1974.

[30] W. E. Ryan, "A Turbo Code Tutorial," New Mexico: New Mexico State University.

[31] D. Rebollo-Monedero, A. Aaron, and B. Girod, "Transforms for High-Rate Distributed Source Coding," in *Proceedings of Asilomar Conference on Signals, Systems and Computers* California, USA: IEEE, 2003.

[32] J. G. Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Communications Letters,* vol. 5, pp. 417-419, 2001.

[33] I. Richardson and Y. Zhao, "Video Encoder Complexity Reduction by Estimating SKIP MODE Distortion," Proceedings of IEE Visual Information. Engineering, 2003.

[34] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive distributed source coding using Low-Density Parity-Check codes," in *Proc. Asilomar Conference on Signals and Systems* Pacific Grove, CA: IEEE, 2005.

[35] A. D. Liveris, X. Zixiang, and C. N. Georghiades, "Joint source-channel coding of binary sources with side information at the decoder using IRA codes," in *Multimedia Signal Processing, 2002 IEEE Workshop on,* , pp. 53-56, 2002.

[36] J. Ascenso and F. Pereira, "Adaptive Hash-Based Side Information Exploitation For Efficient Wyner-Ziv Video Coding," in *IEEE International Conference on Image Processing* San Antonio, USA: IEEE, 2007.

[37] W. Yangli, J. Jechang, and W. Chengke, "An Approach to Side Information Estimation for Wyner-Ziv Video Coding," in *Image and Signal Processing, 2008. CISP '08. Congress on,* pp. 405-410, 2008.

[38] L. Natrio, C. Brites, J. Ascenso, and F. Pereira, "Extrapolating Side Information for Low-Delay Pixel-Domain Distributed Video Coding," in *Int. Workshop on Very Low Bitrate Video Coding*. vol. 3893/2006 Italy, Italy: Springer, pp. 16-21, 2005.

[39] C. Yaacoub, J. Farah, and B. Pesquet-Popescu, "A Genetic Frame Fusion Algorithm For Side Information Enhancement In Wyner-Ziv Video Coding," in *17th European Signal Processing Conference* Glasgow, Scotland: IEEE, 2009.

[40] Eduardo Peixtoto, R. L, d. Queiroz, and D. Mukherje, "On Side Information generation for Wyner-Ziv Video coding," in *XXVI SIMPOSIUM* RIO DE JANEIRO, Brazil 2008.

[41] J. Areia, C.Brites, J. Ascenso, and F. Pereira, "Wyner-Ziv Stereo Video Coding Using a Side Information Fusion Approach," in *IEEE International Workshop on Multimedia Signal Processing* Gerte, Greece: IEEE, 2007.

[42] J. Ascenso, C. Brites, and F. Pereira, "Improving Frame Interpolation with Spatial MotionSmoothing for Pixel Domain Distributed Video Coding," in *5th EURASIP Conference onSpeech and Image Processing, Multimedia Communications and Services* Slovak Republic: IEEE, 2005.

[43] R. Thoma and M. Bierling, "Motion compensating interpolation considering covered and uncovered background," in *Signal Process., Image Commun.* vol. 1, pp. 191–212,1989.

[44] P. Lai-Man and M. Wing-Chung, "A novel four-step search algorithm for fast block motion estimation," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 6, pp. 313-317, 1996.

[45] T. Zahariadis and D. Kalivas, "A Spiral Search Algorithm For Fast Estimation Of Block Motion Vectors," in *Eighth European Signal Processing Conference* 1996, pp. 1079-1082.

[46] M.Ghanbari, "The cross search algorithm for Motion Estimation," *IEEE Transactions on Communications,,* vol. 38, pp. 950-953, July 1990.

[47] Kwon Moon Nam, J.-S. Kim, and R.-H. Park, "A Fast Hierarchical Motion Vector Estimation Algorithm Using Mean Pyramid," *IEEE Transactions on Circuits and Systems for Video technology,,* vol. 5, pp. 344-351, August 1995 1995.

[48] Kyoung Won Lim and J. B. Ra, "Improved Hierarchical Search Block Matching Algorithm by Using Multiple Motion Vector Candidates," *Electronic Letters,* pp. 1771- 1772, October, 1997.

[49] S. Argyropoulos, N. T. Nikolaos, V. Boulgouris, and M. G. Strintzis, "Adaptive Frame Interpolation for Wyner-Ziv Video Coding," in *Multimedia Signal Processing, IEEE 9th Workshop*: IEEE, pp. 159-162, 2007.

[50] A. Trapanese, M. Tagliasacchi, S. Tubaro, J. Ascenso, C. Brites, and F. Pereira, "Exploiting spatial Redundancy in Pixel domain Wyner-Ziv coding," in *Image Processing, IEEE International Conference*, pp. 253-256, 2006.

[51] S. Borchert, R. P. Westerlaken, R. K. Gunnewiek, and a. R. L. Lagendijk, "On extrapolating side information in distributed video coding,," in *26th Picture Coding System,* Lisbon, Portugal.

[52] M. Tagliasacchi and S. Tubaro, "Hash-Based Motion Modeling in Wyner-Ziv Video Coding," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, pp. I-509-I-512, 2007.

[53] R. P. Westerlaken, S. Borchert, R. K. Gunnewiek, and R. L. Lagendijk, "Dependency Channel Modeling for a LDPC-Based Wyner-Ziv Video Compression Scheme," in *Image Processing, 2006 IEEE International Conference on*, pp. 277-280,2006.

[54] R. P. Westerlakena, R. K. Gunnewiekb, and R. Lagendijka, "The Role of the virtual channel in distributed source coding of video," in *ICIP IEEE International Conference image processing*: IEEE, 2005.

[55] D. Baoguo and S. Hong, "Encoder Rate Control for Pixel-Domain Distributed Video Coding without Feedback Channel," in *Multimedia and Ubiquitous Engineering, 2009. MUE '09. Third International Conference on*, , pp. 9-13, 2009

[56] C. Brites and F. Pereira, "Encoder Rate Control for Transform Domain Wyner-Ziv Video Coding," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, pp. II - 5-II - 8, 2007.

[57] C. Brites and F. Pereira, "Correlation Noise Modeling for Efficient Pixel and Transform Domain Wyner-Ziv video coding," *IEEE transactions on Circuit and systems for video technology*, vol. 18, September 2008.

[58] I. Prakash, V. M. Prabhakaran, and R. Kannan, "Towards a theory for video coding using distributed compression principles," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, , pp. II-687-90 vol.3, 2003.

[59] Y. Altunbasak and N. Kamaci, "An analysis of the DCT coefficient distribution with the H.264 video coder," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, pp. iii-177-80 vol.3, 2004.

[60] T. Eude, R. Grisel, H. Cherifi, and R. Debrie, "On the distribution of the DCT coefficients," in *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, pp. V/365-V/368 vol.5, 1994.

[61] Y. Q. Zhang, M. H. Loew, and R. L. Pickholtz, "A methodology for modeling the distributions of medical images and their stochastic properties," *Medical Imaging, IEEE Transactions on*, vol. 9, pp. 376-383, 1990.

[62] J. Ascenso, C. Brites, and F. Pereira, "Motion Compensated Refinement for Low Complexity Pixel Based Distributed Video Coding," in *IEEE International Conference on Advanced Video and Signal Based Surveillance* Como,Italy: IEEE, 2005.

[63] X. Artigas, M. Tagliasacchi, L. Torres, and S. Tubaro, "A Proposal to Suppress the Training Stage in a Coset-Based Distributed Video Codec," in

Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, pp. II-II, 2006.

[64] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: Architecture, Techniques and Evaluation," in *Picture Coding Symposium* Lisbon, Portugal, 2007.

[65] C. Brites, "Advances on distributed video coding," in *Instituto Superior Técnico,* . vol. MSc Lisbon: Technical University of Lisbon, p. 143, 2005.

[66] Z. Xue, "Research and Developments of Distributed Video Coding," in *School of Engineering and Design*. vol. PhD Brunel: Brunel University, p. 163,2009.

[67] P. F. A. Meyer, "Slepian-Wolf turbo coding of video with non-stationary side-information," in *Faculty of engineering*. vol. Master: Delf university of Technologoy, p. 28, 2004.

[68] W.-J. Chien, L. J. Karam, and G. P. Abousleman, "Block-Adaptive wyner-Ziv coding for Transform-domain Distributed video coding," in *32nd IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2007.

[69] J. D. Areia, J. Ascenso, C. Brites, and F. Pereira, "Low Complexity Hybrid Rate Control For Lower Complexity Wyner-Ziv Video Decoding," in *16th European Signal Processing Conference (EUSIPCO 2008),* Lausanne, Switzerland, : EURASIP, 2008.

[70] M. Morb'ee, J. P. Nebot, A. Roca, A. Pizurica, and W. Philips, "Improved pixel-based rate allocation for pixel-domain distributed video coders without feedback channel," in *Acivs 2007* Delft, the Netherlands 2007.

[71] M. Morbee, J. Prades-Nebot, A. Pizurica, and W. Philips, "Rate Allocation Algorithm for Pixel-Domain Distributed Video Coding Without Feedback Channel," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, pp. I-521-I-524, 2007.

[72] J. L. Martinez, G. Fernandez-Escribano, H. Kalva, W. A. R. J. Weerakkody, W. A. C. Fernando, and A. Garrido, "feedback free DVC architecture using machine learning," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, , pp. 1140-1143, 2008.

[73] D. Vaisey and A. Gersho, "Variable block-size image coding," *IEEE Trans. Signal Processing,* vol. 40, pp. 2040–2060, August 1992.

[74] M. Silveira and M. Piedade, "Variable block sized motion segmentation for video coding," in *Circuits and Systems, 1997. ISCAS '97., Proceedings of 1997 IEEE International Symposium on*, pp. 1293-1296 vol.2, 1997.

[75] M. H. Chan, Y. B. Yu, and A. G. Constantinides, "Variable size block matching motion compensation with applications to video coding," *Communications, Speech and Vision, IEE Proceedings I,* vol. 137, pp. 205-212, 1990.

[76] Y. B. Yu, M. H. Chan, and A. G. Constantinides, "Low bit rate video coding using variable block size model," in *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on* pp. 2229-2232 vol.4,1990,.

[77] J. Lee, "Optimal quadtree for variable block size motion estimation," in *Proc. IEEE Int. Conf. Image Processing* Washington, DC: IEEE, 1995, pp. 480–483.

[78] Z. Jiajun, M. Omair Ahmad, and M. N. S. Swamy, "Quadtree structured region-wise motion compensation for video compression," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 9, pp. 808-822, 1999.

[79] P. Strobach, "Quadtree-structured recursive plane decomposition coding of images," *IEEE Trans. Signal Processing,* vol. 39, pp. 1380–1397, June 1991 1991.

[80] "Turbo codes: principles and applications [Book Review]," *Circuits and Devices Magazine, IEEE,* vol. 18, pp. 30-31, 2002.

[81] D. N. Rowitch and L. B. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo (RCPT) codes," *Communications, IEEE Transactions on,* vol. 48, pp. 948-959, 2000.

[82] P. Hoeher, I. Land, and U. Sorger, "Log-likelihood values and monte carlo simulation– some fundamental results," in *Int Symp. on Turbo Codes and Rel. Topics*, Brest, France, pp. 43-46, 2000.

[83] R. Clarke, "Transform Coding of Images," San Diego,USA: Academic Press, 1999.

[84] S. Soni and J. Pillai, "Usage of Nearest Neighborhood, Decision Tree and Bayesian Classification Techniques in Development of Weight Management

Counseling System," in *Emerging Trends in Engineering and Technology, 2008. ICETET '08. First International Conference on*, pp. 691-694, 2008.

[85] A. Djouadi and E. Bouktache, "A fast algorithm for the nearest-neighbor classifier," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, pp. 277-282, 1997.

[86] S. B. Belhaouari, "Fast and Accuracy Control Chart Pattern Recognition using a New cluster-k-Nearest Neighbor," *World Academy of Science, Engineering and Technology*, vol. 49, 2009.

[87] C. Brites, J. Ascenso, J. Quintas Pedro, and F. Pereira, "Evaluating a feedback channel based transform domain Wyner-Ziv video codec," *Signal Processing: Image Communication*, vol. 23, pp. 269-297, 2008.

[88] M. S. Beg, C. Yoong Choon, and T. Ting Fook, "Performance evaluation of error resilient tools for MPEG-4 video transmission over a mobile channel," in *Personal Wireless Communications, 2002 IEEE International Conference on*, , pp. 285-289, 2002.

PUBLICATIONS

1. A.Elamin, Varun Jeoti, Brahim Belhouari "Light Video Compression System Based on Distributed Video Coding". Book title "Visual Information Processing in Wireless Sensor Networks: Technology, Trends and Applications". Publisher, IGI Global, submitted June 2010, accepted for Publication.
2. A.Elamin, Varun Jeoti, Brahim belhouari "Transform Domain Slice Based Distributed Video Coding" Journal of Engineering Science and Technology (JESTEC), ISSN 1823-4690-submitted February 2010, accepted for Publication.
3. A.Elamin, Varun Jeoti, Brahim Belhouari "Best fit models test for the virtual channel In distributed video coding" Journal of Engineering Science and Technology (JESTEC), ISSN 1823-4690-submitted January 2010, accepted for Publication.
4. A.Elamin, Varun Jeoti,Brahim Belhouari "Region Based Adaptive Distributed Video Coding" International Journal of Computer Theory and Engineering, ISSN-1793-8201Vol.3, No.2, Sept.2010.
5. A.Elamin, Varun Jeoti, Brahim Belhouari "A Machine Learning Approach for Rate Allocation In distributed video coding" journal of the Electronics and Telecommunications Research Institute "ETRI"- submitted August 2010, under review.
6. A.Elamin, Varun Jeoti, Brahim Belhouari "Introducing SKIP Mode in Pixel Domain Distributed Video Coding" 6th International Colloquium on Signal Processing & Its Applications. Malaysia, Malacca May,2010
7. A.Elamin, Varun Jeoti, Brahim Belhouari "'Feed Back Channel Usage Reduction In Distributed Video Coding"' The 3rd International Conference on Intelligent & Advanced Systems (ICIAS2010), Malaysia KL June, 2010.
8. A.Elamin, Varun Jeoti, Brahim Belhouari" 'Introducing SKIP Mode in Transform Domain Distributed Video Coding" The 3rd International Conference on Intelligent & Advanced Systems (ICIAS2010), Malaysia KL, June, 2010.