Advanced Distributed Video Coding Techniques

Murat Bilge Badem

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

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

September 2009

Summary

The requirement for multimedia has been growing rapidly where video communication was once considered an entertainment feature, has increasingly become a necessity in modern lifestyle. Mobile video communication and video surveillance are some of the areas that have attracted a considerable amount of attention and notable research efforts have been made for developments. However, current solutions for video communications are far from the ideal video platforms for these technologies mainly due to the high cost and power consumption of the encoder or, in other words, complexity balance of encoder and decoder. In this thesis, we focus on an emerging technology, Distributed Video Coding (DVC), which is ideally suited for such scenarios where the encoder complexity is a critical factor. A number of technical solutions are proposed to the coding framework, in view of improving the performance of DVC towards the conventional video coding techniques.

In this thesis, several approaches have been discussed in order to improve the performance of DVC. First, the progressive nature of side information generation process is enhanced by iterative side information refinement techniques making use of transform coding properties. Then a residual coding scheme is proposed in order to exploit some of the temporal correlation by low complexity operations at the encoder. A non-linear quantisation technique is proposed exploiting the distribution of the residual components. Afterwards, a block based coding technique is proposed for unidirectional DVC and this technique is further enhanced by an improved reconstruction algorithm. Finally, two modified frame coding structures are considered for DVC targeting typical application scenarios. First technique is the interlaced coding which targets high motion videos and low delay requirements; and the second one is the region of activity based coding targeting applications with still cameras. The performances of all proposed techniques are verified and results for all the proposed techniques considerably improved the performance compared to state of the art.

Key words: Video Coding, Distributed Video Coding, Wyner-Ziv Coding, Unidirectional Distributed Video Coding, Transform Coding, Quantisation

Email: m.badem@surrey.ac.uk

WWW: http://www.ee.surrey.ac.uk/CCSR

Acknowledgments

I would like to take this opportunity to thank the people who have contributed to this thesis.

At the first instance, I would like to thank my principle supervisor Dr. Anil Fernando for every thing he has done and his guidance to finish this project with remarkable achievements. Then I would like to thank my co-supervisor Dr. Stewart Worrall, for his guidance and support especially at the beginning. My thanks also go to Prof. Ahmet Kondoz, for enabling me to start a PhD in this area and his guidance and support throughout my thesis.

Then I would like to acknowledge the support of Dr. Hemantha Kodikara Arachchi, Dr. Marta Mrak and my colleagues, Dr. Rajitha Weerakkody, Lasith Yasakethu, Dr. Jose Luis Martinez and Erhan Ekmekcioglu.

I would like to show my appreciation to Adam Sayer, Christopher Hummersone, Edmund Kirby, Mustafa Kondoz, Emre Unver and all other friends for their valuable contributions to this thesis and excellent company.

I would also like to thank the academic and non-academic staff of University of Surrey for everything they have done for me and this project.

Finally, I would especially like to thank my family for their support and encouragement throughout my life.

Contents

List of Figures

List of Tables

Glossary of Terms

AC	Alternating Current
AVC	Advanced Video Coding
BER	Bit Error Rate
CIF	Common Intermediate Format
dB	Decibel
DC	Direct Current
DCT	Discrete Cosine Transform
DISCOVER	Distributed Coding for Video Services
DPCM	Differential Pulse Code Modulation
DSC	Distributed Source Coding
DVC	Distributed Video Coding
DVD	Digital Versatile Disc
fps	Frame Per Second
GOP	Group of Pictures
HDTV	High Definition Television
HVS	Human Visual System
IDCT	Inverse Discrete Cosine Transform
IST	Instituto Superior Tecnico
ITU-T	International Telecommunications Union- Telecommunications Standardization Sector
JPEG	Joint Photographic Experts Group
JVT	Joint Video Team
kbps	Kilobit Per Second
LAPP	Logarithm of the A Posteriori Probability
LDPC	Low Density Parity Check
LLR	Log Likelihood Ratio
MAE	Mean of Absolute Error
MAP	Maximum A Posteriori
MO	Motion Optimisation
MPEG	Motion Picture Experts Group
MSE	Mean Square Error
PAL	Phase Alternate Line
pdf	Probability Density Function
PRISM	Power-efficient, Robust, hIgh-compression, Syndrome-based Multimedia coding
PSNR	Peak-to-Peak Signal-to-Noise Ratio
QCIF	Quarter CIF

QM	Quantisation Matrix
QP	Quantisation Parameter
RCPT	Rate Compatible Punctured Turbo
RD	Rate Distortion
RGB	Red Green Blue
RSC	Recursive Systemetic Convolutional
SAD	Sum of Absolute Differences
SI	Side Information
SISO	Soft-Input Soft-Output
SNR	Signal-to-Noise Ratio
SSIM	Structural Similarity Index
UDVC	Unidirectional DVC
VISNET-II	Networked Audiovisual Media Technologies
VLC	Variable Length Coding
VQM	Video Quality Metric
WZ	Wyner-Ziv

Chapter 1

1 Introduction

1.1 Preamble

Unprecedented growth of consumer appetite for multimedia contents has become the major driving force governing today's dynamics of the consumer electronic market. Due to the ever falling cost, multi-media enabled consumer electronic devices are within the reach of most of the global population and therefore, they are no longer considered luxury devices. Consequent sheer demand for consumer electronic devices, in return, paves the way to new technological innovations and advancements within the consumer electronic domain that reduces the cost even further while including more demanding functionalities to existing devices. While these tremendous technological innovations and growth of consumer demand touches every corner of the consumer electronic domain, multi-functional portable devices have won unexpected attention. The most commonplace example is that of the mobile phone that includes functionalities of audiovisual communication, television receiver, digital camcorder, media player, personal organiser, text/email transceiver, browser and gaming device at present. Inline with these market dynamics, the demand for more cost effective multimedia enabling technologies, such as low complexity video coding algorithms, is imminent.

Video coding technologies have evolved significantly during the past decade, dominated by the work on ISO/IEC MPEG and ITU-T H.26x standards based techniques. These approaches were characterised by a highly complexity video encoder structure and a significantly lower complexity decoder structure as demanded by many popular applications involving one-to-many topologies. Video capturing and encoding was conventionally limited compared to the largely spread listeners to the mostly broadcast type services. However, this concept is increasingly challenged by the emerging and progressively more popular applications involving more encoders than decoders. Security surveillance systems and a number of other remote monitoring applications involving wireless sensor networks promote the use of widely spread encoder networks sharing a limited number of decoders; hence much more encoders than decoders.

Unlike the early days, when attention was only on the functionality and perhaps the robustness of the device, nowadays consumers expects small, stylish designs as well. According to this new trend, it is a tremendous challenge to put ever increasing number of sophisticated functionalities

into a tiny device. Undoubtedly, the processing power, memory capacity and battery capacity must be compromised to achieve required level of miniaturisation. Therefore, the enabling technologies have to be designed in such a way that they are less resource hungry yet functionally competing. The emerging Distributed Video Coding (DVC) concept has the potential of fulfilling the above requirements. DVC marks a dramatic structural change in video coding by shifting the majority of computational complexity, which conventionally resides in the encoder, towards the decoder. The computationally intensive source redundancy exploitation, which is instrumental in achieving large compression ratios, is accordingly performed by the decoder, with this modified computational complexity balance, the DVC encoder retains extremely less expensive operation [1]. Consequently, inexpensive and energy efficient encoder implantations are possible with DVC. Some of the application scenarios are discussed in the following section:

1.1.1 Target Applications

In the literature, the main application requirements where DVC may bring major benefits are identified as:

i. Flexible allocation of the complexity: The DVC approach provides a flexible allocation of the video codec complexity between the encoder and decoder as it shifts some of the complexity from the encoder to the decoder. This requirement includes low complexity encoding and low power needs.

ii. Improved error resilience: As DVC does not contain a prediction loop as with the conventional video coding and decoder estimation is corrected by the encoder, an intrinsic resilience to transmission errors has been claimed [2].

iii. Multi view: In order to exploit the correlation between the cameras in multi view video scenarios, the DVC approach provides a significant architectural benefit since there is not a joint encoder requirement, i.e. the communication between the cameras is not necessary.

iv. Scalability: In the current scalable codecs, there is a predictive approach from lower layers to upper layers letting the encoder to know the decoding results for the previous layers. However, in DVC there is no prediction loop between scalable layers, i.e. layers can be generated by various, different and unknown codecs.

Following application scenarios are listed by DISCOVER (Distributed Coding for Video Services) [3], and the corresponding requirements are given in Table 1.1:

Table 1.1: DVC application scenarios and matching requirements [3]

Application Scenario	Matching Requirements
Wireless Video Cameras	*i*, *ii*
Wireless Low-Power Surveillance	*i*, *ii*, *iii*
Mobile Document Scanner	*i*, *ii*
Video Conferencing with Mobile Devices	*i*, *ii*
Mobile Video Mail	*i*
Disposable Video Cameras	*i*
Visual Sensor Networks	*i*, *ii*, *iii*
Networked Camcorders	*i*, *ii*
Distributed Video Streaming	*i*, *ii*, *iv*
Multiview Image Acquisition	*iii*
Wireless Capsule Endoscopy	*i*

1.1.2 Motivation

DVC has been initiated recently by two different groups ([1] and [4]) early 21st century which is emerged to support the application scenarios mentioned in the previous section. Since then, DVC has attracted the attention of many researchers for low power and low cost video communication. However, at the time this research started in 2006, it was still a relatively untouched area with a challenging nature. Other motivations are:

- The increasing demand on the applications targeted by DVC
- The unique and attractive features of DVC compared to the parallel conventional technologies

1.2 Research Method

1.2.1 Baseline Technology

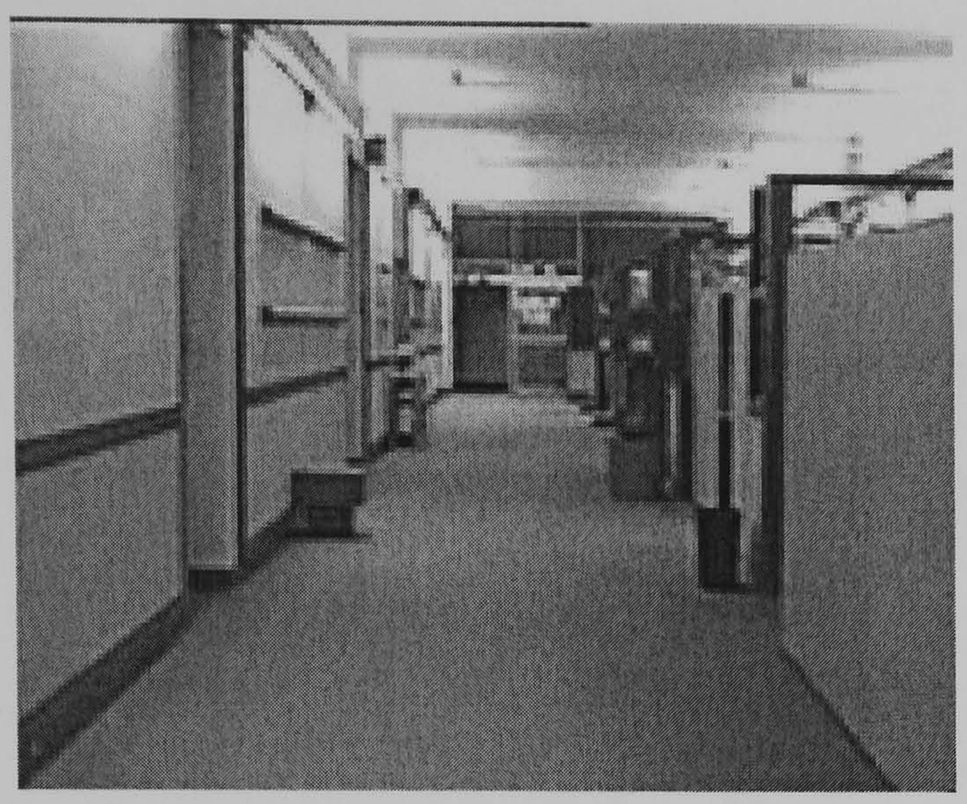
The innovative techniques in this thesis are implemented on top of the baseline DVC codec. As the state of the art evolved, the baseline technology is updated. Due to the better compression efficiency, transform domain coding architecture is selected for the studies and the proposed techniques are all proposed for improving transform domain DVC.

At the beginning of this research, the Instituto Superior Tecnico (IST) DVC codec, which was also chosen as the basic software framework for the European projects DISCOVER [5] and VISNET-II (Networked Audiovisual Media Technologies) [6], is selected as a baseline. Later the VISNET-II DVC codec (an upgraded version of IST DVC codec) is used for the remaining of the studies for this thesis when it became available for this research.

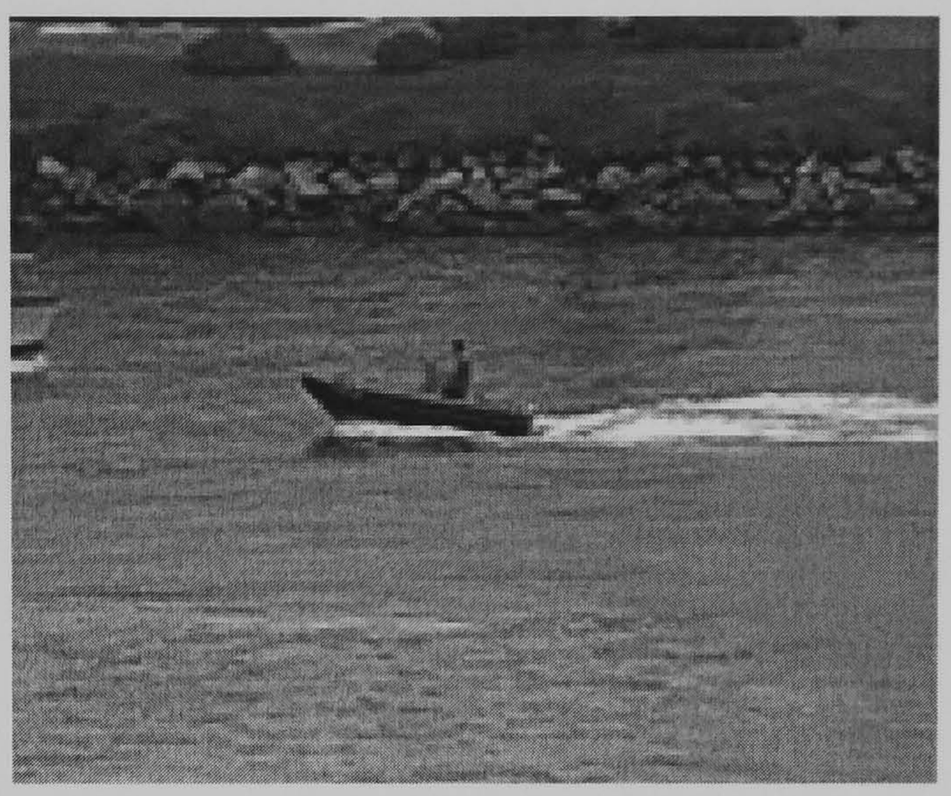
1.2.2 Source Material

Throughout this thesis, a number of standard sequences are used for the tests. These video sequences are short clips which are considered typical for person-to-person video communications. All considered sequences are in YUV 4:2:0 format, which is a usual format for low bit rate coding. In this format, video is represented using one luminance plane (Y) and two chrominance planes (Cb and Cr). The luminance plane corresponds to the black and white information and while the colour information is represented by chrominance planes. The resolution of the luminance plane is twice that of each chrominance planes since the luminance data is perceptually more important than the chrominance data. Following chapter explains video basics in detail.

Four sequences in particular have been used for the majority of the tests carried out in the work here, which are *Hall-Monitor*, *Coastguard*, *Foreman* and *Soccer* sequences. A full list of the sequences used for the test throughout this thesis is given in Appendix A. *Hall-Monitor* and *Coastguard* sequences represent low motion sequences where the former features a stationary surveillance camera in an office and the latter is again a security camera capturing coast view with slight camera movements. However, *Foreman* and *Soccer* sequences represent high motion sequences since they contain considerable amount of motion. First frames from all four sequences are shown in Figure 1-1.



(a) ***Hall-Monitor***



(b) ***Coastguard***

(c) ***Foreman***

(d) ***Soccer***

Figure 1-1: Frames from test videos

1.2.3 Performance Evaluation

The proposed techniques in this thesis are implemented on a baseline codec as discussed above. As the DVC solutions rapidly evolved, the DVC codecs and the test conditions are evolved as well throughout this research. In some tests, original key frames are assumed to be available at the decoder, however later DISCOVER and VISNET-II test conditions are adopted unless the test conditions are different for the bench marks considered in the literature. Therefore, the test conditions vary throughout the thesis and the relevant conditions are documented before describing each simulation results.

In this thesis, the test results are usually presented in the following formats:

- Objective quality vs. bit rate plots: Peak-to-peak Signal-to-Noise Ratio (PSNR) metric is used for all results. However, Structural Similarity Index (SSIM) and Video Quality Metric (VQM) metrics are obtained where necessary.
- Objective quality (PSNR) vs. frame number
- Decoded frames (in order to illustrate the subjective quality)

The details of the above techniques are represented in the following chapter.

1.3 Aim and Objectives

The main aim of this thesis is to improve the performance of transform domain Distributed Video Coding. The specific objectives can be listed as follows:

- Design iterative side information generation algorithms considering transform domain features
- Optimise the quantisation process by considering non-linear techniques
- Design an encoder rate controlling algorithm for unidirectional DVC conserving low encoder complexity
- Optimise the DVC codec through block based coding and interlaced coding

1.4 Novel Work Undertaken

A number of original achievements and contributions have been made to the field of DVC which are supported by several international refereed publications listed in Appendix B. The main novel contributions can be summarised as:

- Iterative side information refinement techniques based on motion interpolation in DC domain
- Improved side information generation technique using modified codec architecture
- Residual coding for transform domain DVC
- Non-linear quantisation algorithm to improve rate distortion performance of DVC
- Block based encoder rate allocation algorithm for Unidirectional DVC
- Region of activity based coding
- Interlaced coding technique for exploiting spatial correlations for intra-coded DVC.

1.4.1 Publications

Some selected international refereed publications related to this thesis are listed below:

Journal Papers:

1. M.B. Badem, M. Mrak and W.A.C. Fernando, "Side information refinement using motion estimation in DC domain for transform-based distributed video coding," *IET Electronics Letters*, Vol: 44, No: 16, pp. 965–966, July 2008.
2. M.B. Badem, W.A.R.J. Weerakkody, W.A.C. Fernando, A.M. Kondoz, "Design of a Non-Linear Quantizer for Transform Domain DVC," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer*, Vol: E92-A, No: 3, pp. 847–852, March 2009.
3. M.B. Badem, W.A.C. Fernando, W.A.R.J. Weerakkody, H. Kodikara Arachchi, A.M. Kondoz, "Transform Domain Unidirectional Distributed Video Coding Using Dynamic Parity Allocation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer*, Vol: E92-A, No: 4, pp.1202–1208, April 2009.
4. M.B. Badem, W.A.C. Fernando, A.M. Kondoz, "Transform Domain Distributed Video Coding with Spatial Correlations," *Multimedia Tools and Applications*, Accepted for publication.

Conference Papers:

5. M.B. Badem, H. Kodikara Arachchi, S.T. Worrall, A.M. Kondoz, "Transform Domain Residual Coding Technique for Distributed Video coding," *Proceedings of Picture Coding Symposium (PCS)*, Lisbon, Portugal, November 2007.
6. M.B. Badem, W.A.C. Fernando, W.A.R.J. Weerakkody, S.L.P. Yasakethu and A.M. Kondoz, "Region-of-Activity Based Coding for Transform Domain DVC," *4th International Conference on Information and Automation for Sustainability (ICIAFS)*, Colombo, Sri Lanka, December 2008.
7. M.B. Badem, W.A.C. Fernando, J.L. Martinez and P. Cuenca, "An Iterative Side Information Refinement Technique for Transform Domain Distributed Video Coding," *Proceedings of IEEE International Conference on Multimedia & Expo (ICME)*, New York, USA, June 2009.

Under Review:

8. M.B. Badem, W.A.C. Fernando, M. Mrak, "Improved side information generation for transform domain distributed video coding," submitted to *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2010.*

1.5 Structure of Thesis

Rest of this thesis is organised as follows:

- Chapter 2: This chapter gives the literature review by explaining; the video coding basics, conventional video coding techniques, Distributed Source Coding (DSC), DVC and the recent work on DVC in detail.

- Chapter 3: In this chapter, some novel side information refinement techniques are proposed. First, a novel DC motion estimation algorithm is presented which forms the basis of the proposed refinement techniques. Then, the proposed techniques are discussed in detail followed by experimental results.

- Chapter 4: In fourth chapter, a novel residual coding technique is proposed for transform domain DVC followed by proposing a dynamic range quantisation technique to optimise the performance of the residual DVC. In the second part of this chapter, a design of a non-linear quantisation technique is presented.

- Chapter 5: A unidirectional DVC solution is proposed in this chapter followed by enhancing the subjective quality by adopting an improved reconstruction algorithm. After discussing both techniques, experimental results are presented.

- Chapter 6: In this chapter, two different frame coding structures have been considered for transform domain DVC. First technique, an intra-coding DVC technique, with interlaced coding is presented. A block based coding technique, based motion of activity levels of each block, is discussed in the second part.

- Chapter 7: Chapter seven contains the overall conclusions for the thesis, and gives suggestions for the future research.

Chapter 2

2 Literature Review

This chapter presents the detailed background for this thesis starting with the video basics and conventional video coding. Following this, Distributed Source Coding principles are introduced, as they form the theoretical basis of Distributed Video Coding. The architectures of currently available Distributed Video Coding frameworks are presented in the following section and recent related works are discussed later.

2.1 Video Coding Basics

2.1.1 Video Signals

Before considering the fundamentals of video coding and compression techniques, it is essential to introduce a few important terminologies and concepts that are used throughout the thesis. Therefore, this section presents an overview of video signals; video colour components, image formats and video quality assessment technologies.

2.1.1.1 Video Colour Components

Raw video signals consist of three colour components: red, green and blue. These components are called RGB signals. As these three colour signals are highly correlated and in order to provide compatibility with black and white video, RGB signals are further processed to generate a new set of signals.

In the popular PAL (phase alternate line) colour system, colour space is represented by YUV: Y (luminance), U and V (chrominance) components. And these components can be calculated using the following equations (where $R'G'B'$ is the gamma corrected RGB) [7]:

$$Y = 0.299R' + 0.587G' + 0.114B' \tag{2.1}$$

$$U = -0.147R' - 0.289G' + 0.436B' = 0.492(B' - Y) \tag{2.2}$$

$$V = 0.615R' - 0.515G' - 0.100B' = 0.877(R' - Y) \tag{2.3}$$

For the digital video signals, the ITU-R Recommendation BT.601 (formerly CCIR-601), has defined YC_bC_r colour space which is very close to YUV space of the PAL system. Y (luminance), Cb and C_r (chrominance) components can be calculated using the following equations [7]:

$$Y = 0.257R' + 0.504G' + 0.098B' + 16 \tag{2.4}$$

$$C_b = -0.148R' - 0.291G' + 0.439B' + 128 \tag{2.5}$$

$$C_r = 0.439R' - 0.368G' - 0.071B' + 128 \tag{2.6}$$

The slight changes to the parameters are limiting the luminance component to the range 16 – 235 and chrominance components to 16 – 240, centred on the grey level 128.

2.1.1.2 Image Formats

The image format may vary significantly based upon the application requirements. For example, for video telephony using mobile phones, small image sizes are preferred for less bandwidth consumption (whilst maintaining acceptable image quality) whereas HDTV (High Definition Television) demands larger frame sizes with improved luminance and chrominance resolutions. Therefore, a series of frame formats have been defined; each format is defined by the spatial resolution of the luminance and chrominance components as well as the temporal resolution. The resolution of the chrominance signal is usually reduced compared to the resolution of the luminance signal due to the nature of human visual perception. This reduction is defined by the *image format* according to the percentage of each chrominance component resolution with respect to the luminance resolution in the horizontal and vertical directions (Table 2.1) [7]:

Table 2.1: Image formats - chrominance resolution as a percentage of luminance resolution

Image Format	Horizontal [%]	Vertical [%]
4:4:4	100	100
4:2:2	50	100
4:2:0	50	50
4:1:1	25	100

CIF:

The Common Interchange Format (CIF) has been defined to enable conversion between different standards, for example between European (625 line, 50 Hz) and North America and Far East (525 line, 60 Hz), for worldwide teleconferencing. CIF defines a video sequence with a resolution of 352 × 288 at 30 frames per second (fps) and an image format of 4:2:0 as illustrated in Figure 2-1[7]:

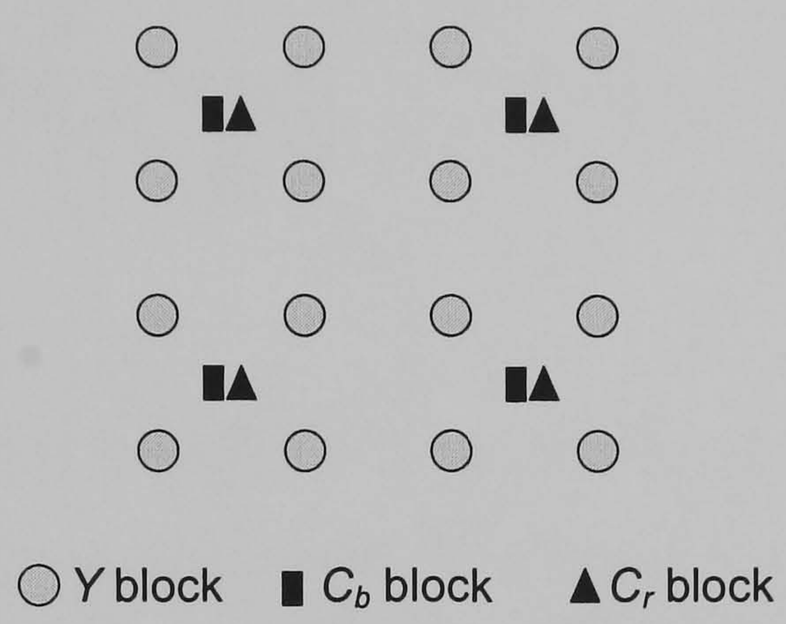


Figure 2-1: Sampling pattern of 4:2:0 image format [7]

Several extensions of CIF have been defined for certain applications. These define the temporal resolutions as 15, 10 and 7.5 fps and spatial resolutions (luminance signal) as listed in the following table:

Table 2.2: Extensions of CIF

Format	Luminance Resolution
SQCIF	128 × 96
QCIF	176 × 144
CIF	352 × 288
4CIF	704 × 576
16CIF	1408 × 1152

2.1.1.3 Video Quality Assessment

Image quality can be measured either subjectively or objectively. For subjective tests, a number of users are required to view and compare a number of video sequences. For each sequence it is necessary to perform a number of tests for different bit rates in order to find an average assessment of the quality. Consequently, subjective tests would require a group of people to spend lots of time for viewing large number of sequences. Without any doubt, this is not a viable

solution and form part of the reason why subjective tests are not often used in video coding research.

The most common and simplest form of objective quality measurement is the ratio of peak-to-peak signal to the root-mean-squared processing noise, which is referred to as peak-to-peak signal-to-noise ratio (PSNR). The equation for PSNR is shown in (2.7):

$$PSNR = 10\log_{10}\left(\frac{255^2}{\frac{1}{N}\sum_{all(i,j)}\left(Y_{ref}(i,j) - Y_{dec}(i,j)\right)^2}\right) \tag{2.7}$$

Where N is the total number of pixels in the image and $Y_{ref}(i,j)$ and $Y_{dec}(i,j)$ are the pixel values of the reference and decoded images. Overall video quality is usually obtained by averaging the PSNR value throughout the sequence either by averaging the PSNR in the Mean Square Error (MSE) or PSNR domain. In this thesis, the latter averaging method is adopted as it is a widely used method in DVC and standard reference software implementations for video codecs (MPEG-4 and H.264).

Table 2.3 gives a rough indication of the PSNR and quality relationship that might be expected from encoding QCIF sequences at reasonably low bit rates. Certainly, these ranges can vary from sequence to sequence.

Table 2.3: Guide to PSNR values (can vary depending on the sequence and encoding options)

PSNR Range (dB)	Quality
< 20	Unacceptable, unintelligible
20 – 25	Subject is perceptible, unacceptable
25 – 28	May be acceptable, but degradation visible
28 – 32	Very little degradation visible
> 32	Good image quality

In some cases there may be large inconsistencies between the PSNR and subjective test results. For example, block errors may not have a significant effect on PSNR if it is due to a single bit error. However, subjects can perceive this as an annoying artefact.

In this study, where necessary, two other metrics have been used in addition to PSNR: Structural-Similarity-Based Image Quality Assessment (SSIM) and Video Quality Metric (VQM). The SSIM metric is based on the fact that the human visual system is highly adapted for extracting structural information from a scene [8]. The equation for SSIM is shown in (2.8):

$$SSIM = \frac{(2\overline{xy} + C_1)(2\sigma_{xy} + C_2)}{(\overline{x}^2 + \overline{y}^2 + C_1)(\sigma_x{}^2 + \sigma_y{}^2 + C_2)} \tag{2.8}$$

Where, $x = \{x_i \mid i = 1,2,\ldots,N\}$ is the original signal and $y = \{y_i \mid i = 1,2,\ldots,N\}$ is the distorted signal. $\overline{x}$, $\overline{y}$, $\sigma_x{}^2$, $\sigma_y{}^2$ and $\sigma_{xy}{}^2$ are the mean of x, mean of y, the variance of x, the variance of y and the covariance of x and y respectively. C_1 and C_2 are constants. The SSIM metric takes values between 0 and 1. The higher the SSIM index the better the quality, therefore 1 corresponds to the highest quality. An 11 × 11 circular-symmetric Gaussian weighting function is used where the mean and variances are obtained locally within this 11 × 11 sliding window. The overall index of the image is the average of all the quality indexes of the image.

The VQM metric is the combination of the perceptual effects of video impairments including global noise, blurring, jerky/unnatural motion, block distortion and colour distortion into a single metric. More information on the VQM evaluation techniques can be found in [9] and [10].

2.1.2 Video Compression

Transmission of uncompressed video requires a high bandwidth and large storage capacities, which can be unrealistic and costly. Therefore, video coding is necessary for compressing the video whilst allowing reversible conversion of data, requiring fewer bits and more efficient transmission. Figure 2-2 illustrates the comparison of uncompressed and compressed frames of *Foreman* sequence. Both frames are in QCIF (176 × 144) 4:2:0 image format. The compressed frame is decoded using the H.264/AVC codec. It should be noted that, although the frame has been compressed to less than 20 kbits from 304 kbits, the drop in the image quality is negligible (> 37 dB).

(a) Uncompressed Frame (> 300 kbits) **(b) Compressed Frame (19.12 kbits)**

Figure 2-2: A comparison of uncompressed and compressed frames of *Foreman* Sequence

Considering the above, video coding aims to exploit the redundancies together with knowledge of the Human Visual System (HVS), in order to achieve compression. The redundancies within a video sequence can be identified as spatial and temporal correlations. Only the spatial correlations can be exploited for coding still images. This is called intra-frame coding. If the temporal correlations are exploited in addition to spatial correlations, it is called inter-frame coding. Inter-frame predictive coding is a key coding principle, which is widely used in standard video codecs. Video compression is based on three fundamental approaches:

- Spatial redundancy reduction: exploiting correlations within a frame
- Temporal redundancy reduction: exploiting correlations between successive frames
- Entropy coding: reducing redundancies between compressed data symbols

2.1.2.1 Spatial Redundancies

Exploiting spatial redundancies usually utilises predictive and transform coding:

- Predictive Coding

Some spatial redundancies are exploited by predicting the values of each pixel or pixel group using previously coded information, and coding the error. This technique is called Differential Pulse Code Modulation (DPCM). The highest contribution to this process comes from the neighbouring pixels due to high spatial correlations. The neighbouring pixels can be either from the same frame or from the adjacent frames. The prediction error is calculated and coded by

comparing the predicted picture with the original picture at the encoder and at the decoder the received signal is added to the prediction. Note that, this is a lossless process.

- Transform Coding

In transform coding, the pixel data is mapped into a transform domain prior to data reduction. In most scenes, image energy is concentrated in low frequency regions, i.e. in a few transform coefficients. Insignificant coefficients can be discarded by quantising the coefficients so that data compression can be achieved. This is a lossy coding technique since some of the coefficients (low significant coefficients are totally discarded) can not be reconstructed.

- Quantisation

The human eye responds differently to different spatial frequency coefficients i.e. if the magnitude of higher frequency coefficients is below a certain threshold the human eye will not detect it. Quantisation is a process that attempts to determine what information can be safely discarded without a significant loss in visual fidelity. In other words, in the transform domain the energy is unevenly distributed and quantisation exploits this property by using different quantisation levels for each band. This is a lossy process as the reconstructed coefficients contain quantisation noise.

2.1.2.2 Temporal Redundancies

Exploiting temporal redundancies usually makes use of motion estimation algorithm.

- Motion Estimation

Motion Estimation techniques are based on block matching each block and the best matching block in the reference frame. First, frames are divided into blocks with $m \times n$ dimensions. Then this block is compared against all blocks (within a pre-defined search window) of the same size in the reference frame. The search window is usually larger than the block size and the searched block is centred on the search window. The best matching block is chosen by considering several approaches including Mean Squared Error (MSE) and Mean of Absolute Error (MAE). The motion vector is the spatial displacement between the searched block and the best match. Once all the blocks are searched within the reference frame, all motion vectors form a motion vector map. This map is then used to generate the predicted frame by compensating with the reference frame.

2.1.2.3 Variable Length Coding

This technique reduces the redundant information in a bit stream independently of the spatial and temporal correlations. Variable Length Coding (VLC) uses short code words for highly probable values and long code words for the less probable ones. Huffman coding and arithmetic coding are the most common VLC techniques and are widely employed in standard video codecs.

2.2 Conventional Video Coding Techniques

Video coding techniques that do not follow the distributed source coding principles are called *conventional* in this thesis. In this section, a historical view of conventional video coding techniques followed by an overview of conventional encoder and decoder architectures are given.

2.2.1 History of Video Coding

Initial attempts at video coding for communications systems go back to the 1960s when an analogue video phone system was tried out. Formal and organised video coding groups, the International Telecommunications Union- Telecommunications Standardization Sector (ITU-T) and the Joint Photographic Experts Group (JPEG), made had standardisation attempts in late 1980s. Figure 2-3 illustrates the evolution of video coding standards starting from 1984 [7].

H.261 standard codec used a combination of inter-frame Differential Pulse Code Modulation (DPCM) and Discrete Cosine Transform (DCT). It was initially directed at video coding at 384 kbit/s, later extended to other bit rates (multiplies of 64 kbit/s).

The Motion Picture Experts Group (MPEG) investigated coding technologies for video storage, such as CD-ROM in the early 1990s. The first generation of MPEG standard, MPEG-1, used H.261 as a starting point. MPEG-1 was optimised for non-interlaced video at 1.2-1.5 Mbit/s rates. Following this, a new generation of standards emerged for coding interlaced video at higher bit rates (4-9 Mbit/s) called MPEG-2. MPEG-2 had a significant impact in many digital video applications such as terrestrial broadcasting, satellite TV, cable TV and digital versatile disc (DVD). Later, MPEG-2 was adopted by the ITU-T under the generic name of H.262 for telecommunications. H.262/MPEG-2 featured scalability, an important feature for video networking applications such as video on demand and multicasting.

After numerous developments on MPEG-1 and MPEG-2, the MPEG group started working on a codec (MPEG-4) operating at very low bit rates (64 kbit/s or less). ITU-T, in parallel, carried out some work on a new standard for similar target applications and a new codec named H.263 was devised. With the improved compression efficiency over the years, the evolutions of this standard were named as H.263+ and H.263++. In 1997, two groups, the ITU-T and the ISO/IEC MPEG group came together and formed a Joint Video Team (JVT) in order to create a single video coding standard: H.26L. Later, this codec was published jointly as Part 10 of MPEG-4 and ITU-T Recommendation H.264 [11] and was called Advanced Video Coding (AVC).

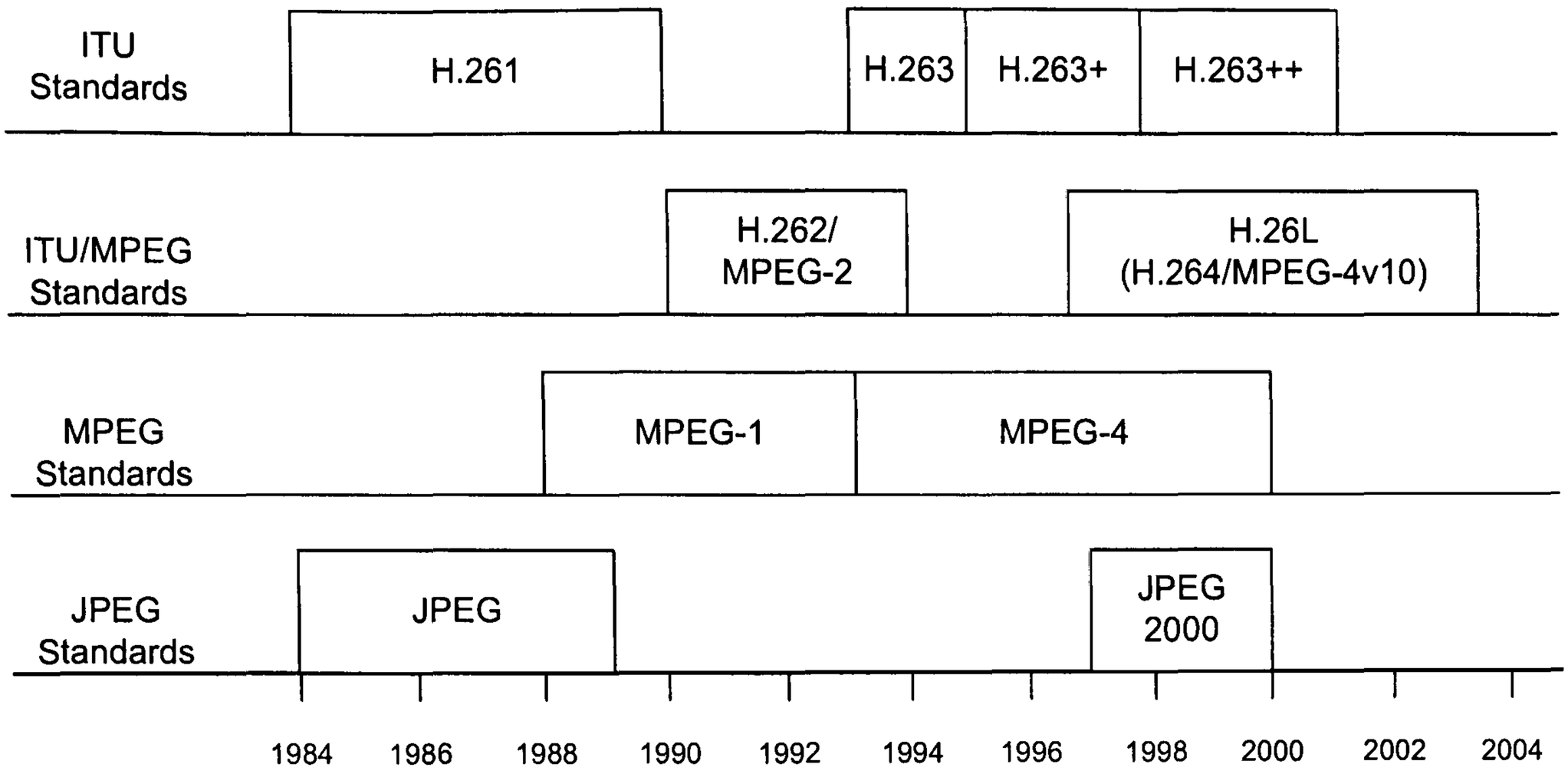


Figure 2-3: Evolution of video coding standards [7]

2.2.2 Conventional Video Encoder

To achieve the requirements of random access, a set of pictures can be defined to form a Group of Pictures (GOP). In general, there are three basic types of frames in a GOP: Intra pictures (I), Unidirectional (forward) predicted pictures (P) and Bidirectional predicted pictures (B). Intra frame coding (I) exploits the spatial correlations within the same frame, whereas P and B coding exploits the temporal correlations with the reference frame(s). Each GOP includes one I frame and several P and B frames, a typical GOP structure is illustrated in Figure 2-4.

Once the predicted frame (either intra or inter prediction) is obtained, the original frame to be encoded is compared with the prediction signal. The prediction error (residual signal) is then transformed (DCT), quantised and entropy coded with additional data (including motion vectors and quantisation step sizes). A replica of the decoded frame is obtained by inverse quantisation, inverse DCT transform and adding the residual signal in order to get the predicted frame. This encoding procedure is shown in Figure 2-5.

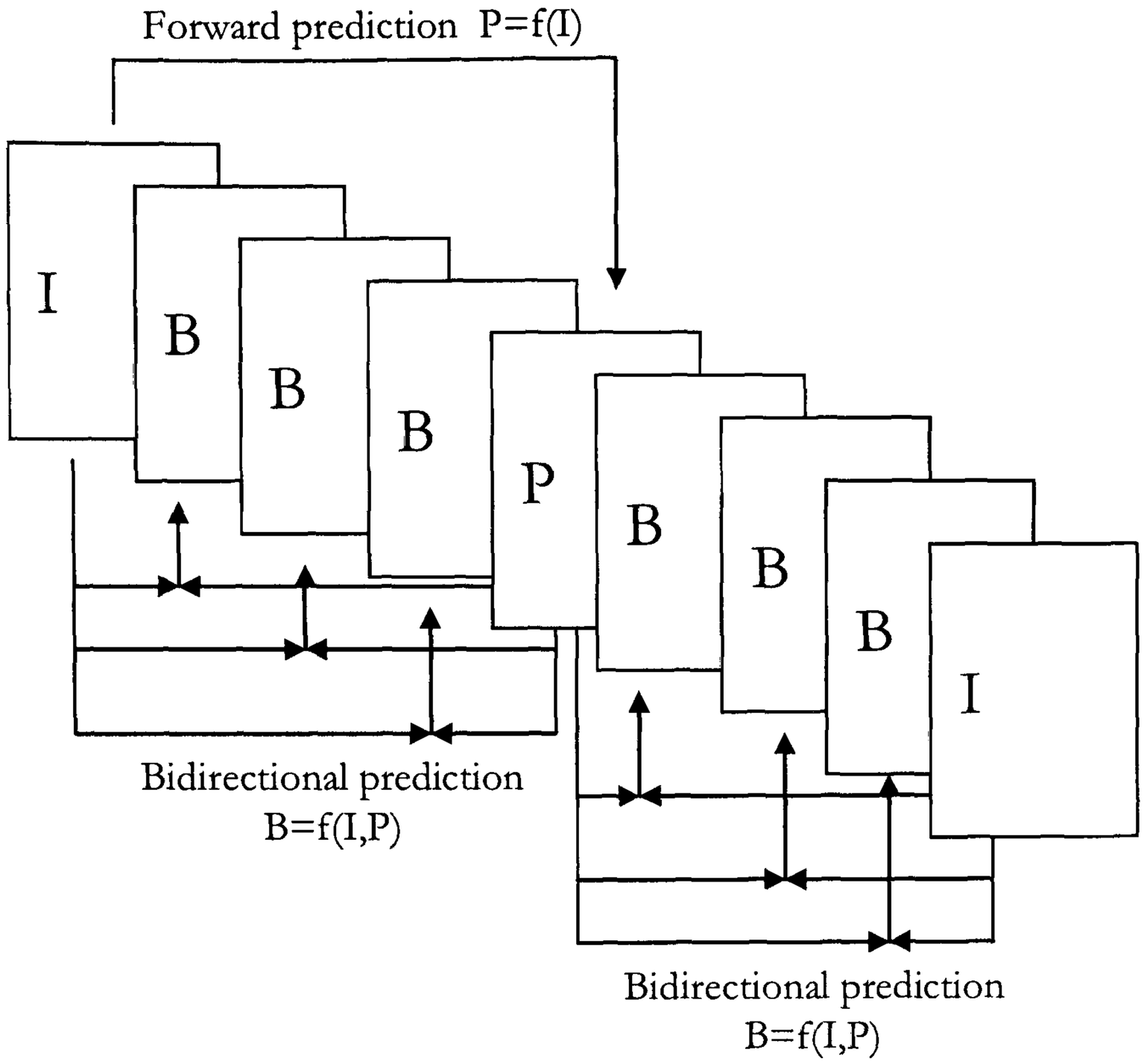


Figure 2-4: A typical GOP structure for video coding

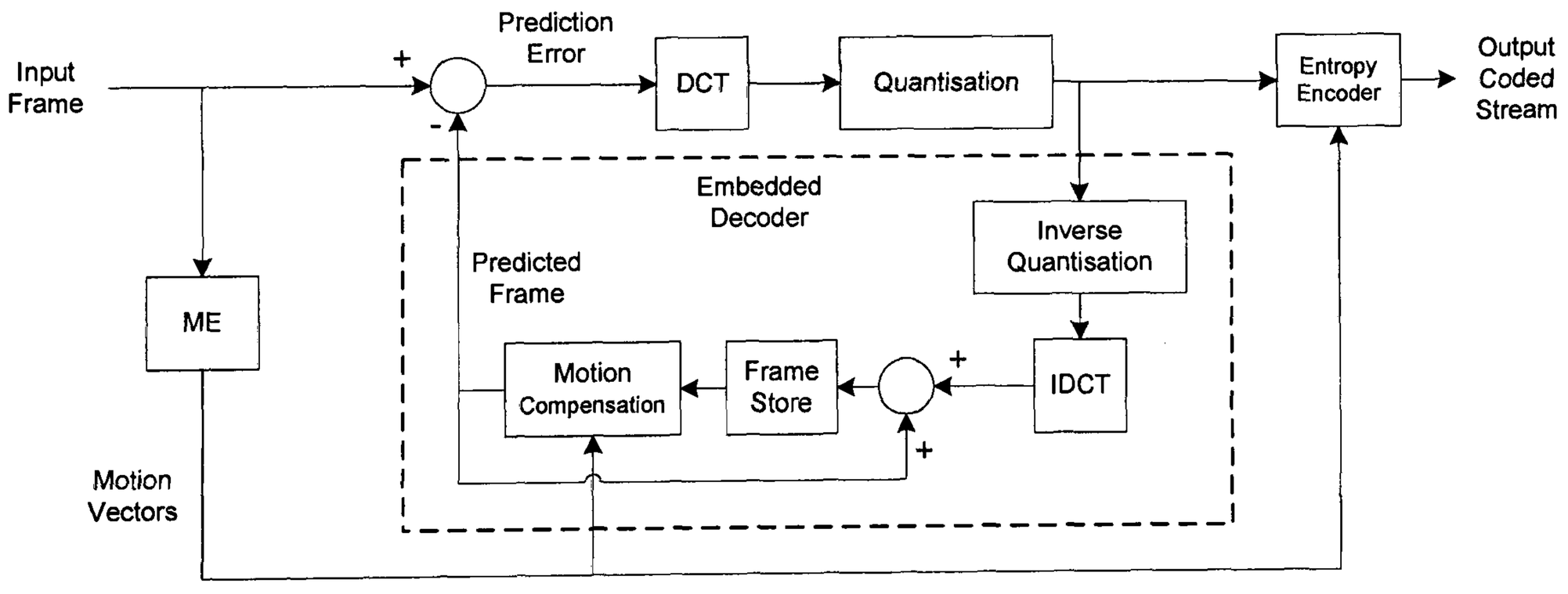


Figure 2-5: A typical conventional video encoder

2.2.2.1 Transform Coding

In conventional video coding the Discrete Cosine Transform (DCT) is the most commonly used transform coding technique. The DCT function for the integer 4×4 block based transform is described in this section (as defined in H.264/AVC [11]). A two-dimensional matrix transform is achieved through two one-dimensional transforms in both the horizontal and vertical directions for each row and column of the 4×4 sample matrix from the source picture. The following 4×4 transform matrix is utilised for this purpose [11]:

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \tag{2.9}$$

Matrix multiplication is performed on all non-overlapping 4×4 blocks of the frame. This operation can be calculated using 16 bit arithmetic using only addition, subtraction and shifts. Each 4×4 block of pixels is transformed into a 4×4 block of DCT coefficients with each coefficient representing some degree of spatial correlation within the pixel block. The significance of the information carried in each coefficient varies in terms of image energy. The lower frequency components, which correspond to the top-left coefficients of the DCT block (Figure 2-6), are more significant when considering the HVS, as the human eye is more sensitive to low frequency spatial information components.

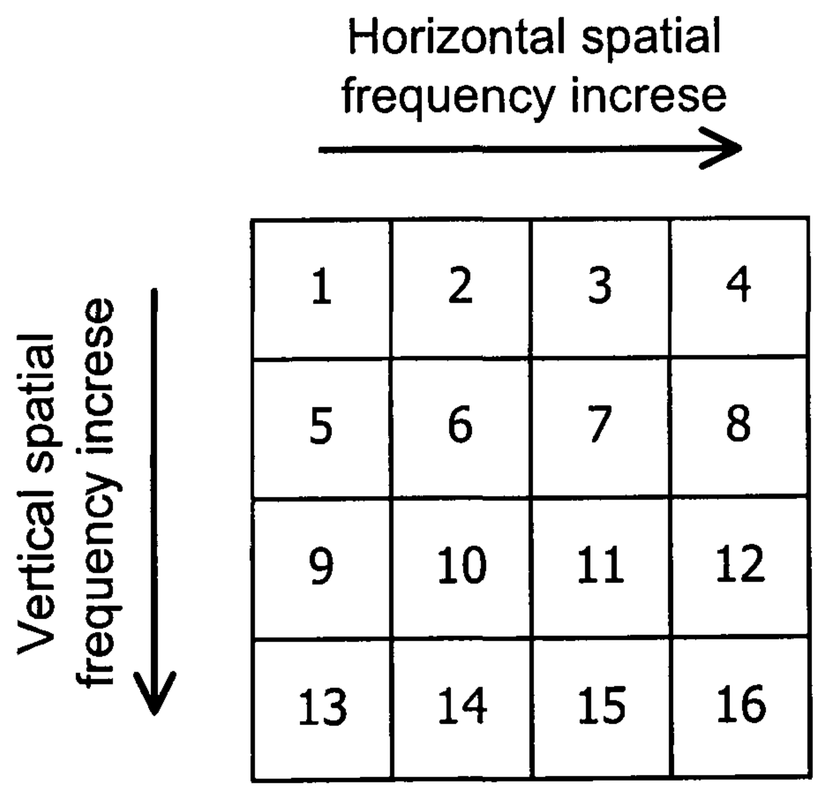


Figure 2-6: Transform coefficients

The inverse DCT is then performed at the decoder using the following inverse transform matrix (as defined in H.264/AVC [11]).

$$\tilde{H}_{inv} = \begin{bmatrix} 1 & 1 & 1 & 1/2 \\ 1 & 1/2 & -1 & -1 \\ 1 & -1/2 & -1 & 1 \\ 1 & -1 & 1 & -1/2 \end{bmatrix} \tag{2.10}$$

2.2.2.2 Motion Estimation

Motion vectors of each block are encoded together with the prediction error. A block matching algorithm is often utilised by partitioning the frame into blocks of $M \times N$. The block matching algorithm is generally performed only on the luminance data to reduce its complexity. As shown in Figure 2-7, a search is performed for each block in the current frame within an area in the frame called the *search window*. The block matching algorithm assumes that all pixels within the current block go in the same direction over the search window. The best match for each block of the current frame is found from the reference frame providing the minimum distortion. There are several criteria for this evaluation including MSE and MAE:

$$MSE(i,j) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left(X_C(m,n) - X_R(m+i,n+j) \right)^2 \tag{2.11}$$

$$MAE(i,j) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left| X_C(m,n) - X_R(m+i,n+j) \right| \tag{2.12}$$

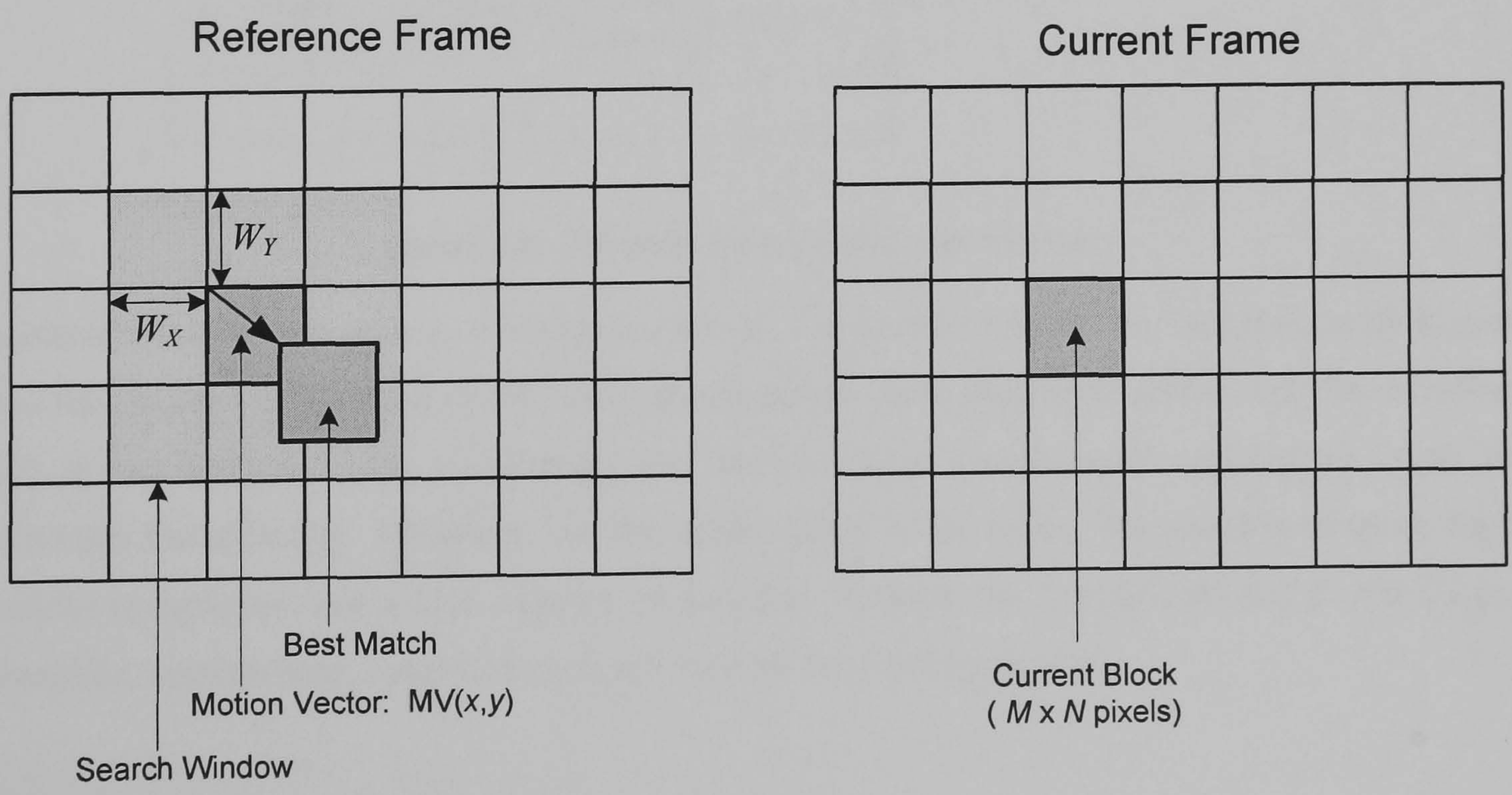


Figure 2-7: Motion estimation

Where i and j are the search points within the search window: $-W_X < i < W_X$ and $-W_Y < j < W_Y$. X_C (m,n) and X_R (m,n) represent the current block and reference blocks respectively and $(m+i,n+j)$ is the location of the candidate pixel in the reference corresponding to the location (m,n) in the current frame.

The (i,j) couple that gives the minimum distortion according to equations (2.11) or (2.12), represents the shift distance in horizontal and vertical directions and hence the corresponding motion vector.

2.2.3 Conventional Video Decoder

A typical decoder architecture is shown in Figure 2-8. The encoded bit stream is entropy decoded, inverse transformed and inverse quantised to obtain the prediction error (residual signal). The prediction error is then added to the predicted signal, which is motion compensated using the motion vectors and the reference frames (previously obtained), to produce the output frame. The decoded frame is stored for the future frames.

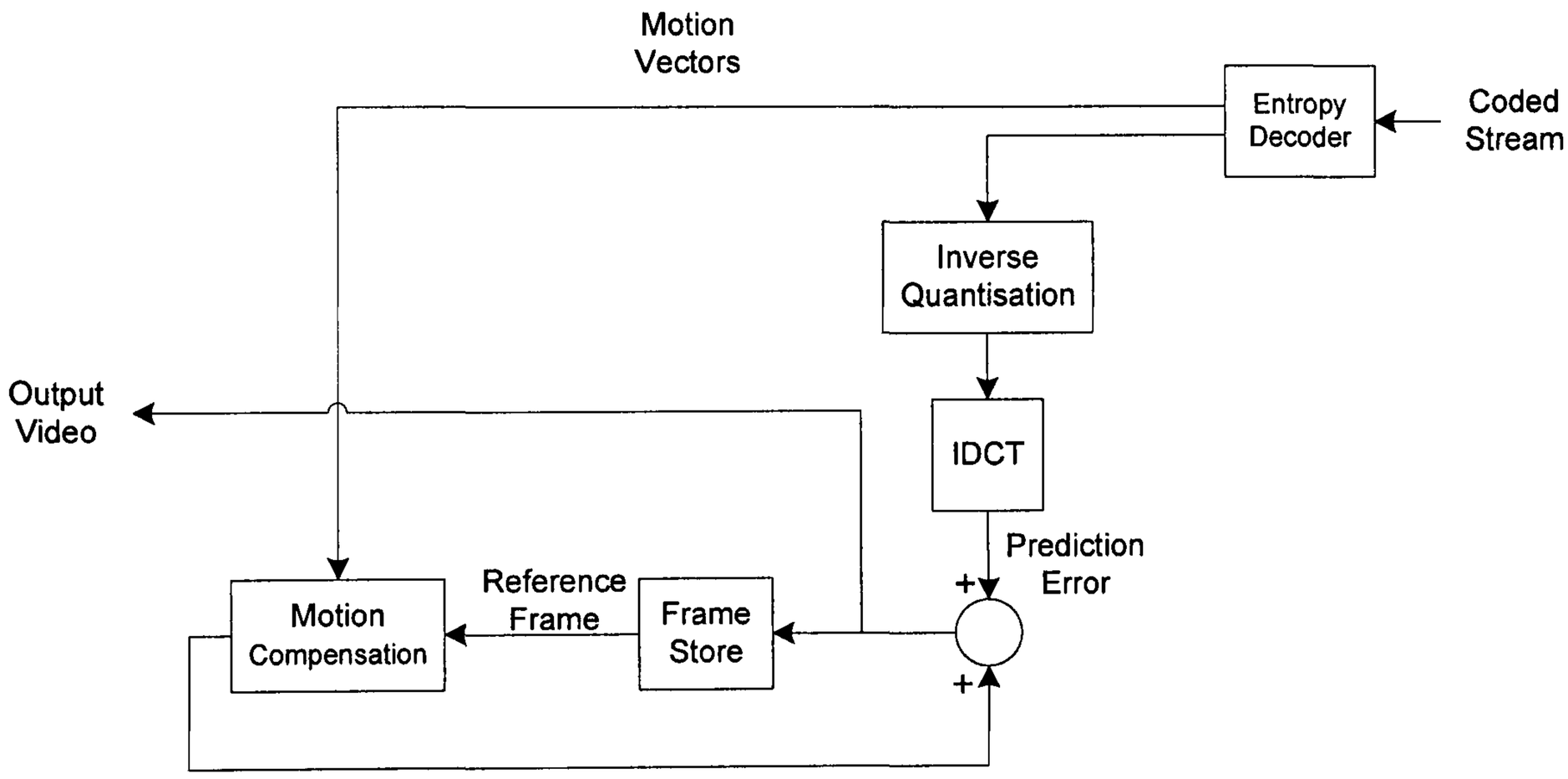


Figure 2-8: A typical conventional video decoder

In conventional video coding, as discussed above, the encoder complexity is significantly higher than the decoder complexity due to motion estimation, intra-prediction search and the decoding path at the encoder. This encoder-decoder balance suits one-to-many applications, such as television broadcasting. However, for the applications listed in the introduction chapter, high encoder complexity and a high number of encoders increase the overall cost and power usage. Therefore, conventional video coding is not suitable for such applications.

2.3 Distributed Source Coding

DVC is an adaptation of the theoretical framework of distributed source coding (DSC) set by the Slepian-Wolf theorem [12] and the Wyner-Ziv [13] theorem for video coding. The DSC concept deviates from the conventional source coding paradigm in the context of the dependency of encoding statistically correlated sources. In the conventional approaches, statistically correlated sources are jointly encoded and jointly decoded for perfect reconstruction of the information stream at the decoder. The main goal of the conventional video coding is keeping the encoding rate, R, minimum in order to perfectly reconstruct two statistically dependent sequences X and Y (Figure 2-9) using a joint decoder. This encoding rate, given in (2.13), is the joint entropy of X and Y. DSC, in contrast, attempts to carry out independent encoding of statistically dependant sources that can be jointly decoded. The theoretical information limitations inherited by so-called independent encoding have been reported by Slepian and Wolf, as discussed in the next section (section 2.3.1).

$$R = H(X,Y) \tag{2.13}$$

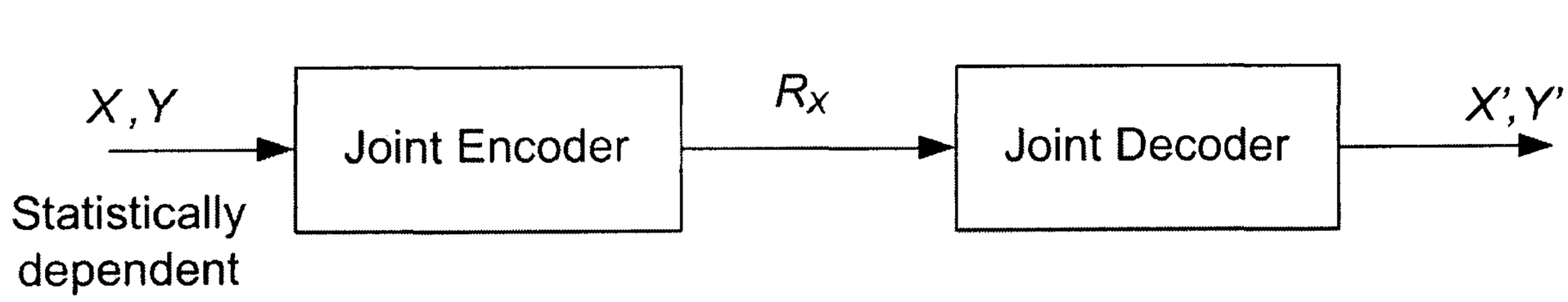


Figure 2-9: Conventional coding of two statistically dependent random sequences

2.3.1 Slepian-Wolf Theorem

Assume that X and Y are two statistically dependent discrete random sequences and are independently and identically distributed (i.i.d.). In separate encoding and decoding of the two sequences, X and Y, error-free (no coding error) reconstruction can be achieved with rates:

$$R_X \geq H(X) \tag{2.14}$$

$$R_Y \geq H(Y) \tag{2.15}$$

where R_X and R_Y are encoding rates and $H(X)$ and $H(Y)$ are the entropies of X and Y. Now consider the case that these sequences are separately encoded but jointly decoded, exploiting the correlation between them as illustrated in Figure 2-10. The total bit rate, $R = R_X + R_Y$ is naturally greater than $H(X)+H(Y)$. However, in the 1970s Slepian and Wolf studied Distributed Coding and provided the following results [12]. The possible rate combinations of R_X and R_Y for reconstruction of X and Y with an arbitrarily small error probability are shown below:

$$R_X \geq H(X \mid Y) \tag{2.16}$$

$$R_Y \geq H(Y \mid X) \tag{2.17}$$

$$R_X + R_Y \geq H(X,Y) \tag{2.18}$$

Where $H(X|Y)$ and $H(Y|X)$ are the conditional entropies and $H(X,Y)$ is the joint entropy of X and Y. According to the Slepian-Wolf theorem, if the total overall bit rate exceeds the summation of individual bit rates $R_X + R_Y$ and the conditional entropy between X and Y is lower than the summation, then the independently encoded streams could be jointly decoded with an arbitrarily small bit error probability. Thus, the lower bound to the bit rate is set to be equal to the summation of individual bit rates.

Thus it is concluded that the independent encoding of statistically dependent sequences does not impose any theoretical loss of compression efficiency when compared with more established approaches of joint encoding that are utilised in conventional video coding techniques.

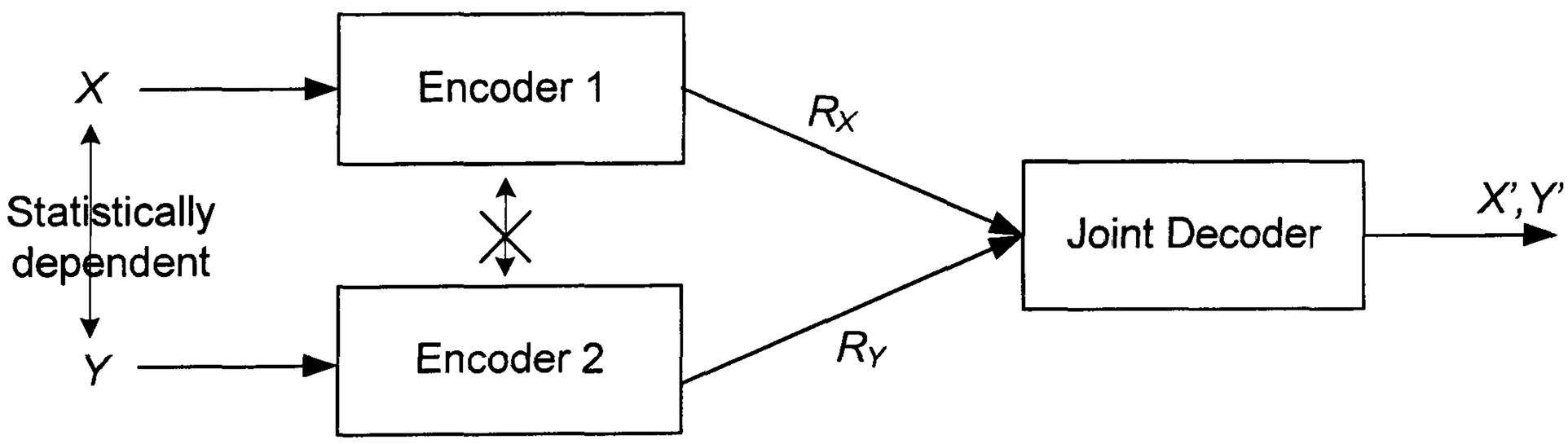


Figure 2-10: Distributed coding of two statistically dependent discrete random sequences

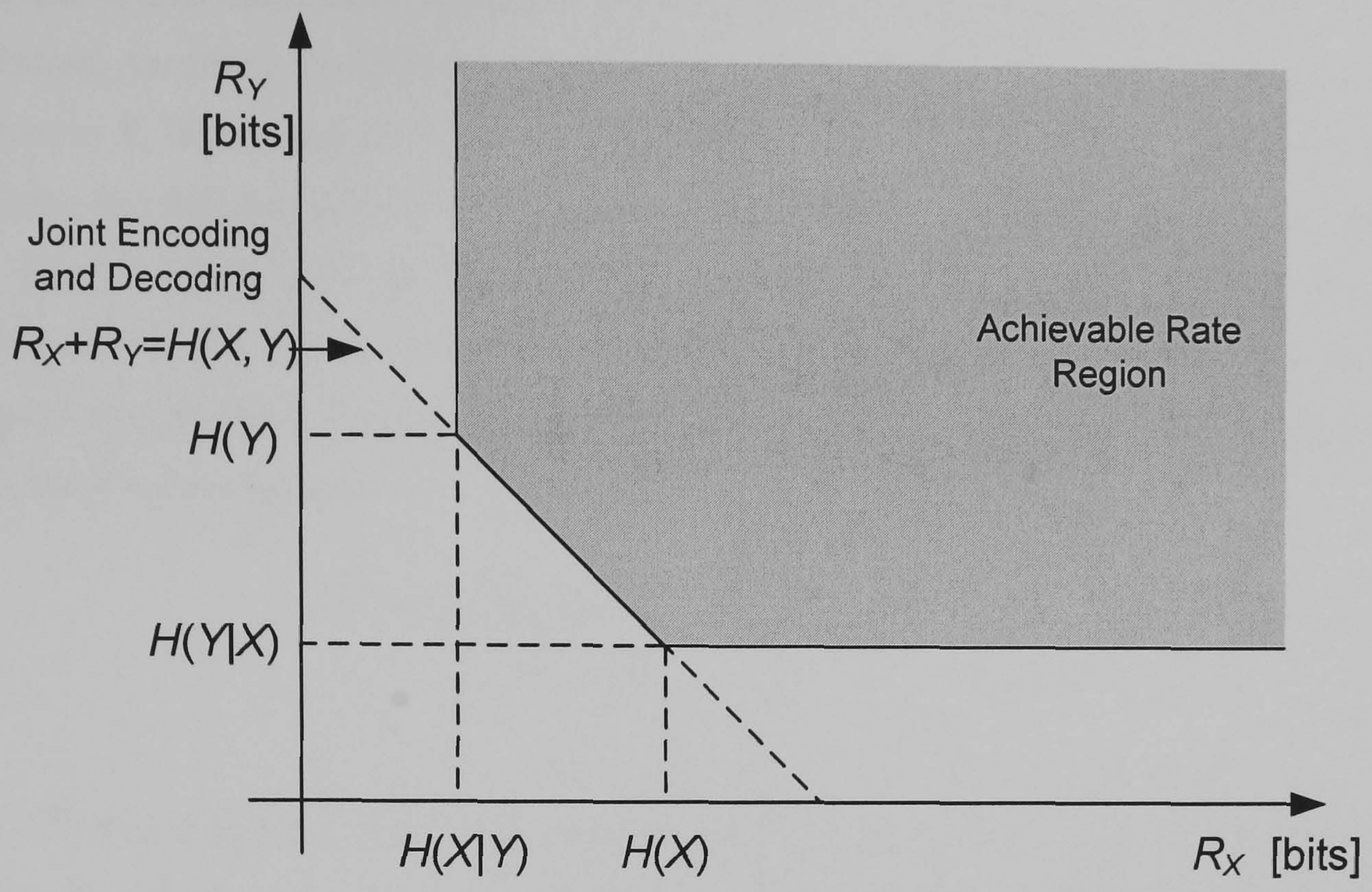


Figure 2-11: Achievable rate region following the Slepian-Wolf theorem

Figure 2-11 illustrates the achievable rate region for the distributed coding of two statically dependent i.i.d. sources, X and Y for the recovery of information with an arbitrarily small error probability according to the Slepian-Wolf theorem [12]. In Figure 2-11, the vertical, horizontal and diagonal lines, correspond to Equations (2.16), (2.17) and (2.18) respectively, represent the lower bounds for the achievable rate combinations of R_X and R_Y.

2.3.2 Wyner-Ziv Theorem

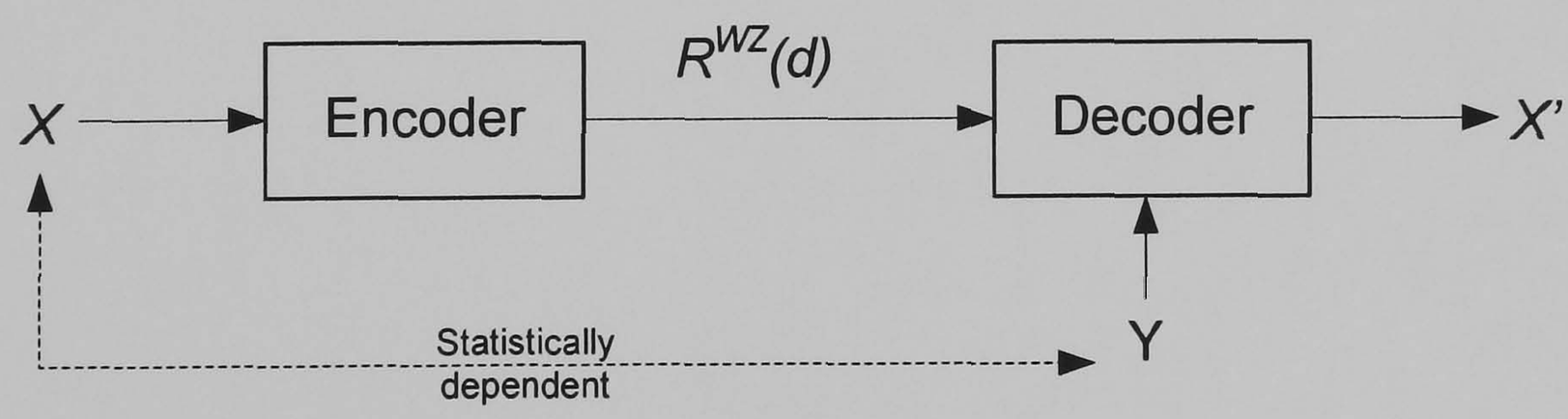


Figure 2-12: Lossy compression with decoder side information

The Wyner-Ziv theorem [13] refers to a particular case of the Slepian-Wolf theorem [12] which is commonly known as the lossy compression scenario with side information at the decoder. This scenario is shown in Figure 2-12. This concept assumes a finite acceptable distortion level d,

between the source information X and the corresponding decoded output X'; hence the term *lossy compression*. Assuming the information content Y is available at the decoder, which is statistically dependant to X, Wyner and Ziv attempted to quantify the minimum bit rate to be passed between the encoder and decoder, termed $R^{WZ}(d)$ for achieving the finite distortion d between the input and output. The Wyner-Ziv theorem states 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. The Wyner-Ziv theorem reads:

$$R^{WZ}(d) \geq R_{X|Y}(d), \qquad d \geq 0 \tag{2.19}$$

Where $R^{WZ}(d)$ is called the *Wyner-Ziv minimum encoding rate* (for X) and $R_{X|Y}(d)$ represents the minimum rate necessary to encode X when Y is simultaneously available at the encoder and decoder for the same average distortion d.

When d approaches zero, i.e. when no distortion exists, the Wyner-Ziv theorem falls back to the Slepian-Wolf result, i.e. $R^{WZ}(0) \geq R_{X|Y}(0)$. This means that it is possible to reconstruct the sequence W with an arbitrarily small error probability, even when the correlation between W and the side information is only exploited at the decoder.

2.4 Distributed Video Coding

Having realised the applicability of distributed source coding principles to video signals, initial studies into Distributed Video Coding started in 2002. Since then, Distributed Video Coding has been an attractive alternative to conventional video coding for a number of applications where lightweight encoders are needed, such as security surveillance and remote monitoring. A low computational complexity encoding concept has been achieved by moving redundancy exploitation to the decoder side. This section presents an overview of the basic DVC frame work on which this study is based. The first part, section 2.4.1, compares the complexity comparison of conventional video coding and the DVC.

2.4.1 Complexity Balance of Video Codecs

In video compression, the most computationally complex operations are the spatial and temporal redundancy exploitations. In conventional approaches to video coding with joint encoding and joint decoding, such as ISO/IEC MPEG-X and ITU-T H.26x, these complex operations are carried out at the encoder. Conversely, the decoder complexity is much lower, allowing it to be computationally inexpensive. The complexity of the encoder in conventional video coding structures is expected to be 5 to 10 times higher than the decoder complexity [14]. Figure 2-13 illustrates a symbolic representation of this complexity balance in conventional video codecs. These architectures are well suited to traditional one-to-many video applications including television broadcasting and multicasting. An example scenario for video communication with one encoder and many decoders is given in Figure 2-14. The complexity balance of conventional video coding is justified in these scenarios because lower decoder complexity reduces production costs.

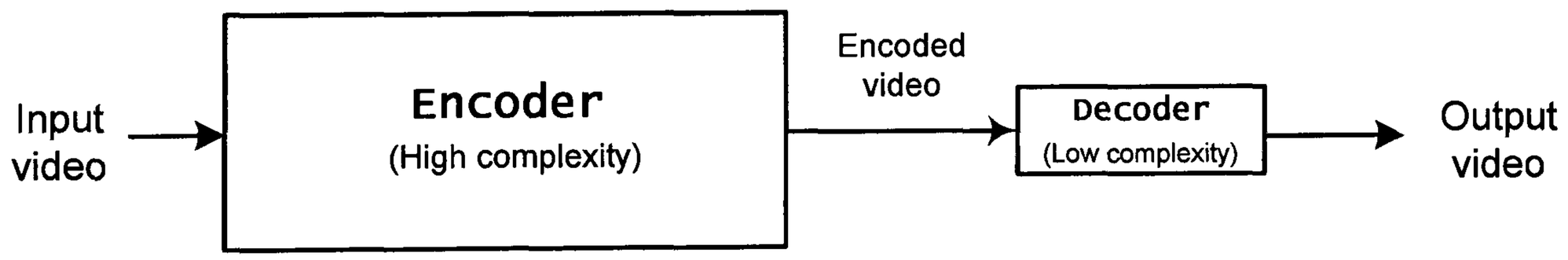


Figure 2-13: Complexity balance of the encoder and decoder of conventional video coding

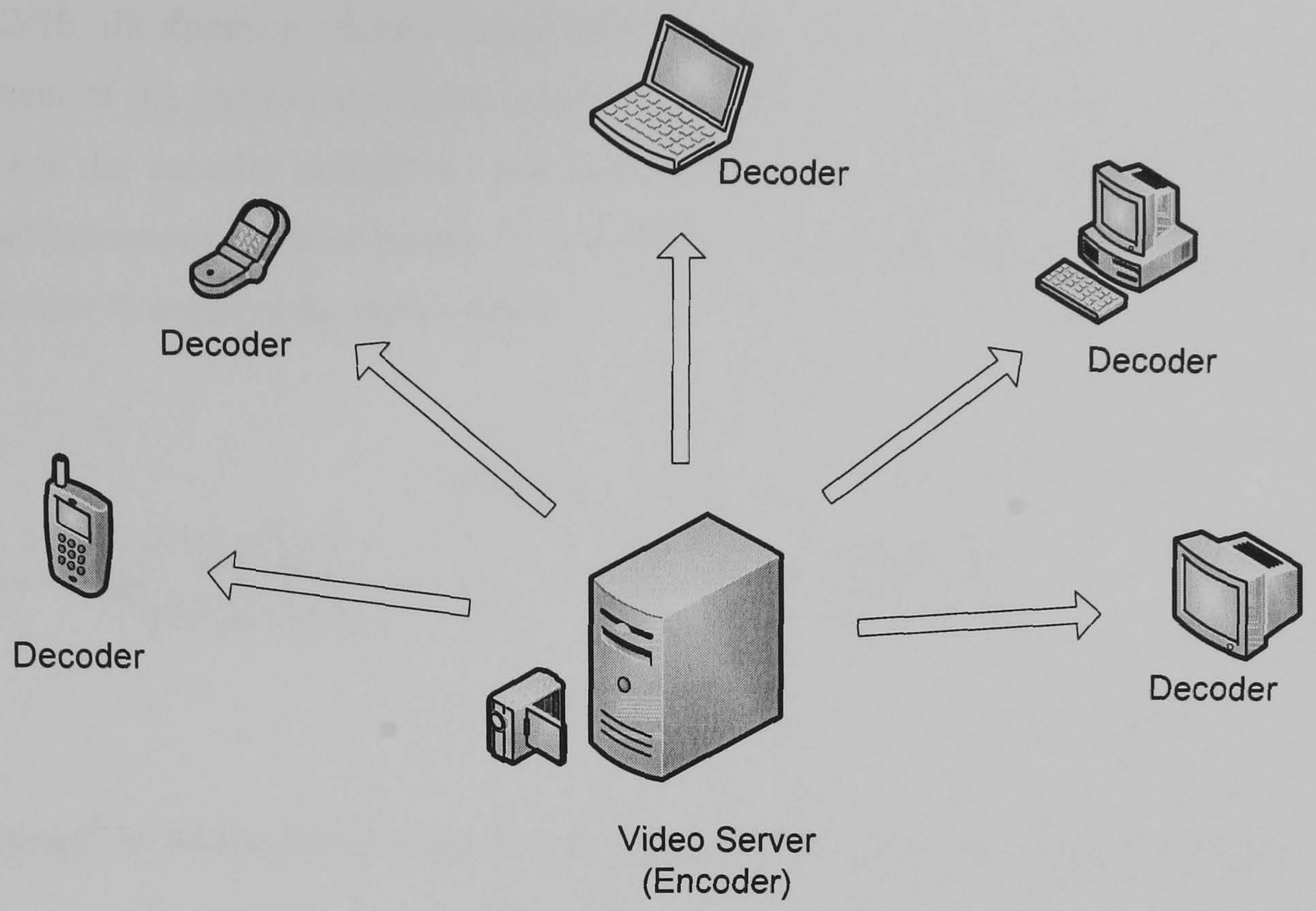


Figure 2-14: Typical one-to-many application scenario

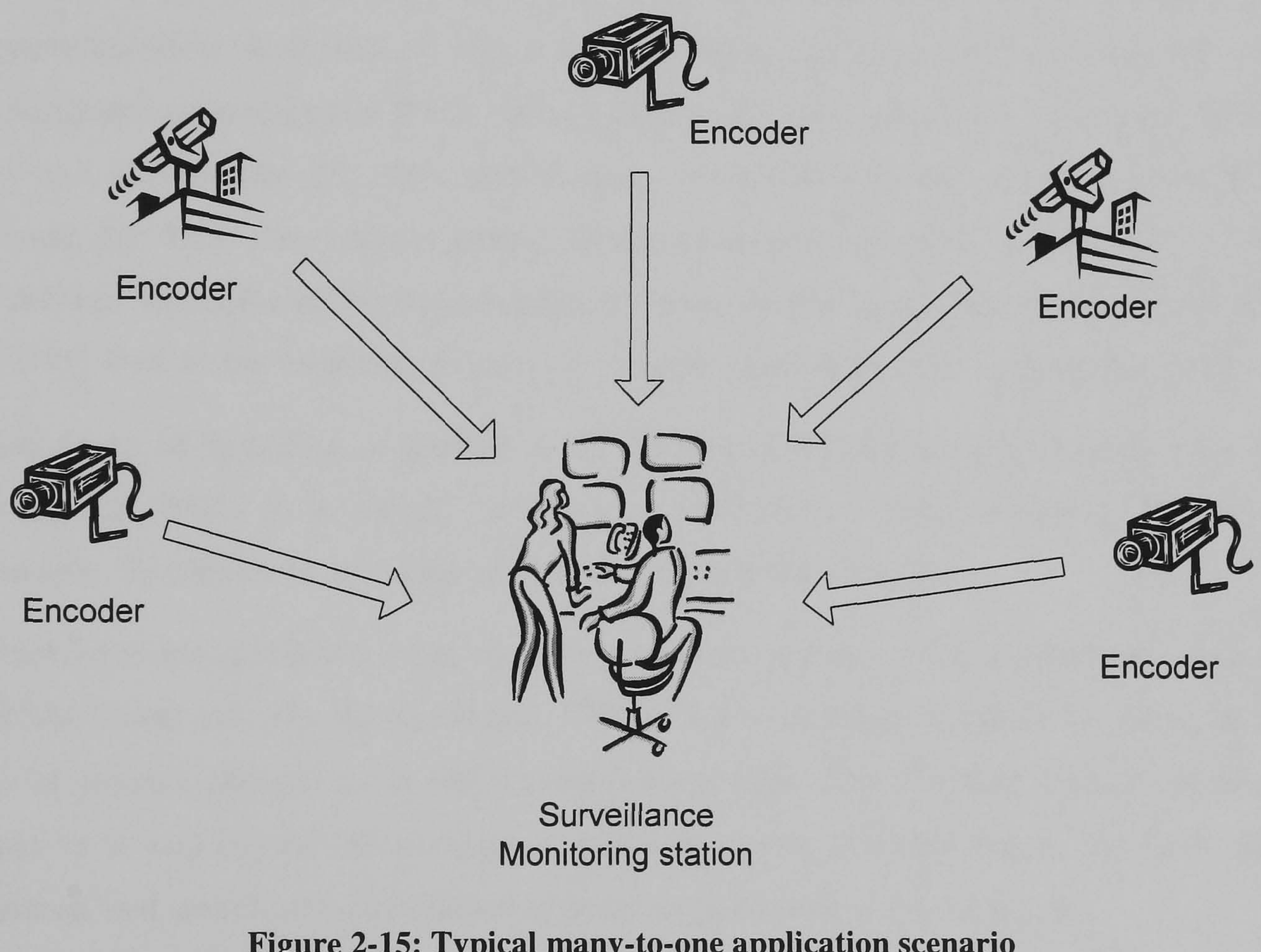


Figure 2-15: Typical many-to-one application scenario

Many-to-one type applications, such as video surveillance, in contrast to the above scenario, require low cost video encoder architectures. An example security surveillance scenario is illustrated in Figure 2-15. In this scenario, a large number of video sensors capture, encode and upload the video to the central monitoring station.

Figure 2-16 illustrates a shifted complexity balance that would fulfil the low complexity requirement of the one-to-many applications. The observations made of basic DVC codecs have shown that the encoder complexity could be reduced to negligible levels. However, several additional operations, such as transform coding and encoder rate control, have been added to the DVC encoder to improve the performance.

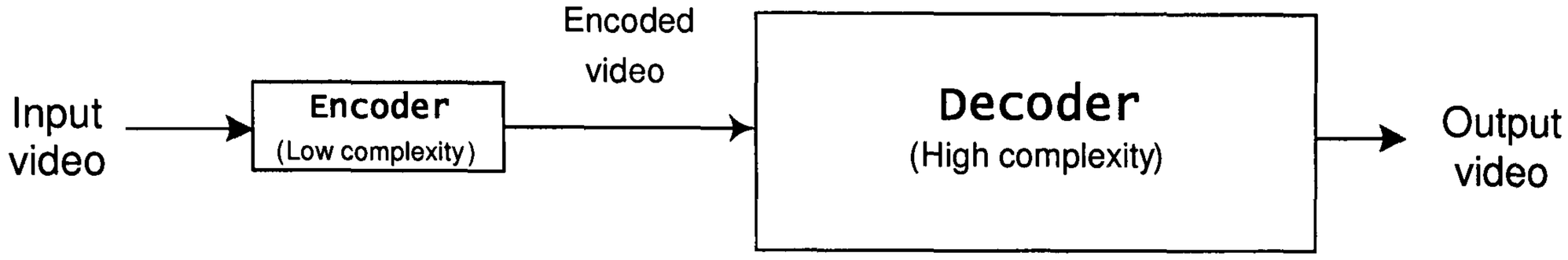


Figure 2-16: Complexity balance of the encoder and decoder of Distributed Video Coding

2.4.2 Relevant DVC Codec Solutions

Two initial independent proposals have been made with regard to the development of a DVC framework based on the distributed source coding principals. Firstly, the University of Stanford in 2002, suggested a pixel domain DVC, introducing two different types of frames called Wyner-Ziv frames and Key frames [1]. This architecture, commonly referred to as the Stanford DVC framework, has been very popular among recent researchers due to its low encoder complexity and a detailed discussion of the implementation. Later, the Stanford research group extended their initial DVC work to the transform domain [15] and hash-based motion compensation [16].

The University of Berkeley, in parallel to the Stanford DVC framework, presented the second framework in 2002. It is usually referred as PRISM or "Power-efficient, Robust, hIgh-compression, Syndrome-based Multimedia coding" in the literature [4].

The Stanford framework has a lower encoder complexity and easier implementation, compared to the PRISM framework. On the other hand, PRISM is a more robust architecture, and uses no key frames or reverse channel as in the Stanford framework. The Stanford framework has been followed in a majority of the subsequent research works, possibly due to the more flexible architecture, and several structural improvements are presented in the literature.

Later in 2005, Instituto Superior Tecnico de Lisboa (IST) developed its own DVC codec following the Standford framework in both the pixel domain [17] and transform domain [18]. Both architectures outperform the Stanford framework. The IST DVC codec was chosen as the basic software framework for the European projects DISCOVER and VISNET-II (a European Network of Excellence). Therefore the IST codec and later the VISNET-II codec— an improved version of IST codec—are selected as a basis for this thesis.

2.4.2.1 Stanford Codec Architecture

The Stanford group has proposed several DVC solutions for the pixel domain including [1] and [19] in 2002 and 2003 respectively. In 2004, transform domain DVC solutions, [15] and [16] were also proposed.

In the Stanford approach, the input video sequence is divided into Wyner-Ziv (WZ) frames and key frames such that even frames are WZ frames and odd frames are key frames. Key frames are assumed to be losslessly available at the decoder [1]. However, later in [19], a more flexible approach was presented, where the number of WZ frames between key frames may vary and key frames are encoded using the H.263+ standard codec. Each WZ frame is encoded independently from the key frames and other even frames. In the transform domain solution, [15], each DCT band is separately encoded. To decode a WZ frame, the side information (prediction of that frame) is generated through motion compensated temporal interpolation techniques using the key frames. Note that, the techniques presented in [1], [15] and [19] all are based on an intra-frame-encoder–inter-frame-decoder structure.

Later, another solution was presented by Aaron et al. in [16]. In this solution, aside from the bit stream produced for WZ frames (as explained above), supplementary information is generated and transmitted to help the decoder in the motion extrapolation task. Aaron et al. considered this solution as a near intra-frame-encoding–inter-frame-decoding solution because the additional computation and memory usage for the supplementary information is much lower compared to traditional inter-frame coding.

2.4.3 Pixel Domain DVC Codec Architecture

In the literature, DVC architectures are presented as Wyner-Ziv video coding solutions since they are based on the Wyner-Ziv theorem, i.e. lossy source coding with side information at the decoder. In these techniques, some frames of the video sequence are encoded using traditional video coding solutions in order to achieve the $H(Y)$ rate compression proposed by the Wyner-Ziv Theorem. On the other hand, other frames are encoded using the Slepian-Wolf theorem in order to achieve the rate $H(X|Y)$. Therefore, the objective of DVC is to approach the rate point $(H(X|Y),H(Y))$ (see Figure 2-11, page 24). Channel coding techniques and quantisation are used in order to achieve this target.

Figure 2-17 illustrates the architecture of the pixel domain DVC utilising turbo coding in the Slepian-Wolf codec. Important modules are identified at the encoder as: quantiser, bit plane extractor, Slepian-Wolf encoder, and key frame encoding; at the decoder as: Slepian-Wolf decoder, side information generator, reconstructor, and key frame decoder. There is a feedback

channel as in the original Stanford DVC framework, which is used to communicate dynamic parity request messages from the decoder to encoder buffer.

At the encoder, the input video sequence is divided into Wyner-Ziv frames and key frames. Wyner-Ziv frames are quantised and turbo encoded whereas key frames are traditionally encoded. At the decoder, the quantised bit stream is decoded through joint source-channel decoding with the help of the side information bit stream. Side information is again used together with the decoded bit stream in the reconstruction module in order to estimate the transform coefficients. Finally, the output frame *X'* is obtained by inverse transforming the reconstructed coefficients.

Each component is explained in the following subsections in detail.

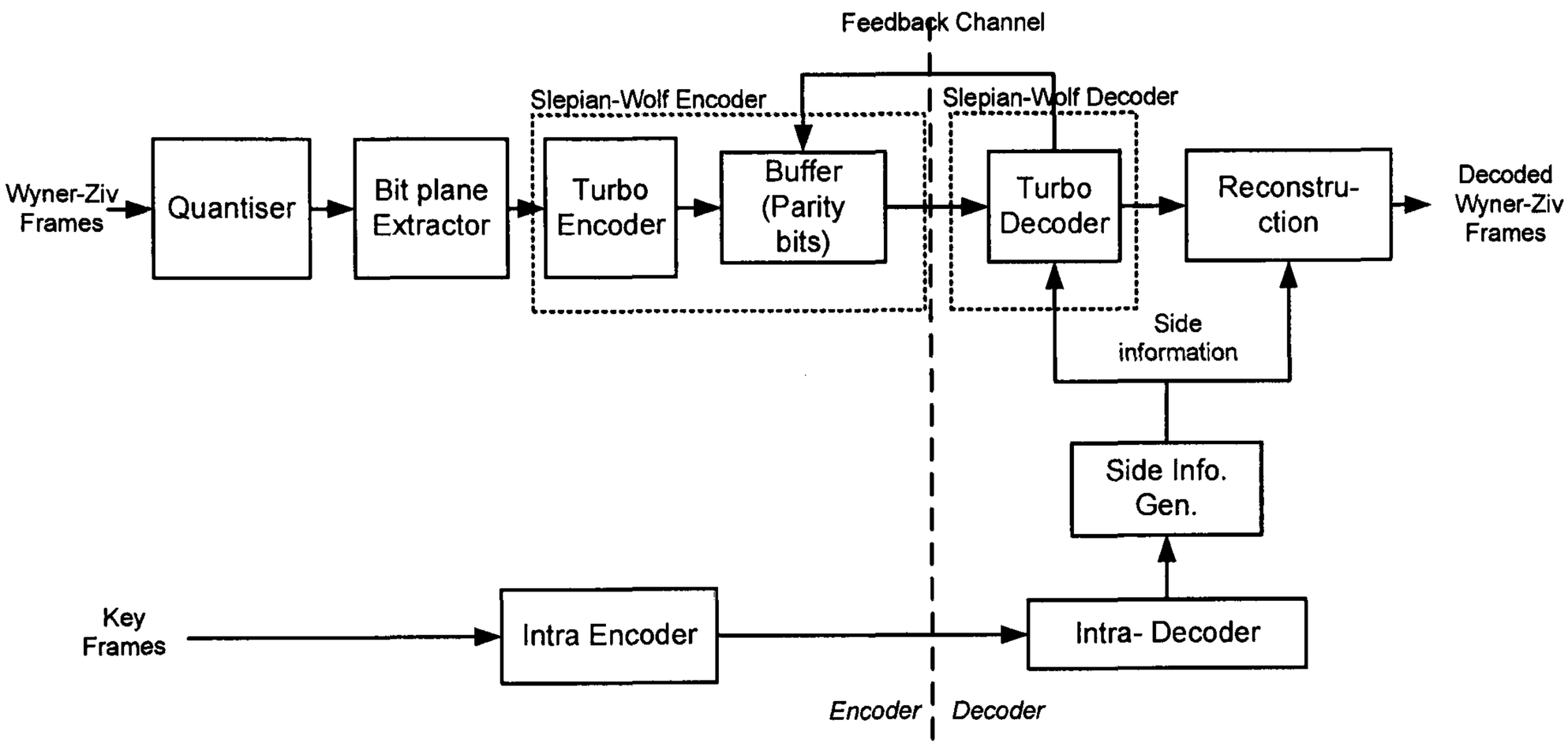


Figure 2-17: Pixel domain DVC architecture

2.4.3.1 Quantisation and Bit Plane Extraction

Quantisation is an important function in pixel domain DVC as well as in any signal coding algorithm. In pixel domain DVC, quantisation is the process of limiting the number of discrete levels. Each pixel is represented by 8 bits, resulting in 256 quantisation levels in the unquantised video signal. A series of quantisation bins is generally used to define the quantisation function. This is illustrated in Figure 2-18, where *M*, the quantisation parameter, determines the number of bins.

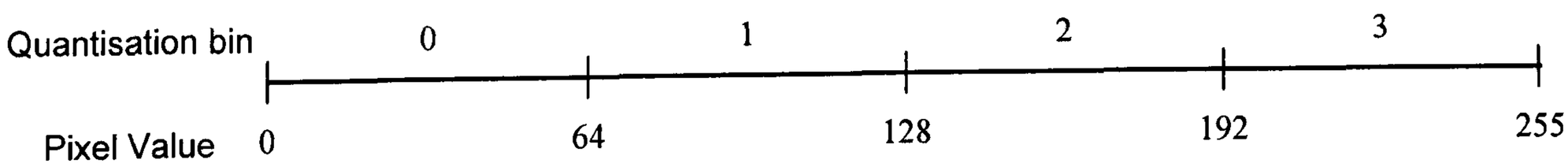


Figure 2-18: Quantisation bins for $2^M = 4$

Quantisation is an irreversibly lossy operation since the number of discrete levels is limited (quantisation noise). However, the reconstruction function of the DVC codec partially compensates for the quantisation noise.

The bit plane extractor groups bits of the same significance to prepare the bit stream for Slepian-Wolf coding. In the bit plane extraction process, bits with the same significance form the bit planes. The most significant bits form the first bit plane, the second most significant bits form the second bit plane and so on for all bits representing each pixel.

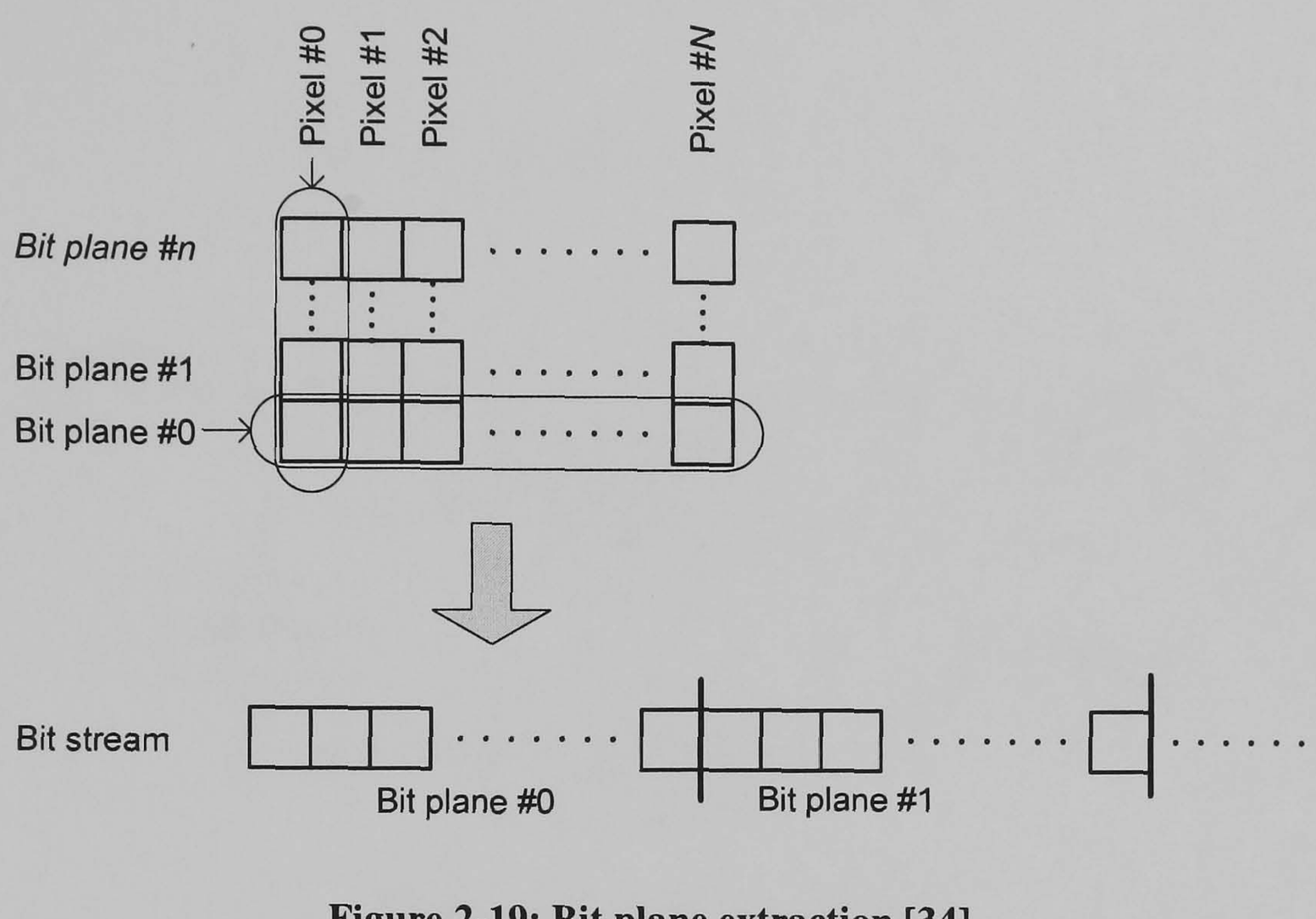


Figure 2-19: Bit plane extraction [34]

In pixel domain DVC, the quantisation and bit plane extraction processes are directly linked in the implementation. Therefore, quantisation is achieved by discarding the desired number of bit planes of lowest importance (see Figure 2-19).

2.4.3.2 Slepian-Wolf Encoder

In DVC, channel coding techniques are utilised to correct the side information with respect to the original information whilst reducing the transmission rate. The first steps towards a practical design used syndromes [20] or embedded trellis codes [21]. Later, turbo codes were tested by Garcia-Frias and Zhao [22], [23]; and Barjcsy and Mitran [24], [25] using binary random sequences. Aaron and Girod [26] significantly improved the performance of turbo coding by using Gaussian sequences. Low Density Parity Check (LDPC) codes have also been applied to distributed source coding by Liveris et al. [27], [28]. Recently designed Slepian-Wolf codecs, based on turbo codes or LDPC codes [29], [30], [31], [32], show performances closer to the theoretical bounds explained previously in this chapter. Turbo coding, followed by the majority of the researchers, is explained below.

Turbo Encoder

Turbo coding was proposed in 1993 by Berrou et al. [33] for channel coding in communication systems. Due to its simplicity of implementation and efficient performance, turbo coding has been one of the most popular options for Slepian-Wolf coding. The turbo encoder consists of two parallel convolutional encoders separated by an interleaver (see Figure 2-20). The structure of the convolutional encoder is defined by the generator polynomial $G(D)$ of the form given in (2.20), where $g_1(D)$ and $g_2(D)$ are the feed-forward and feedback polynomials respectively and D denotes the delay.

$$G(D) = [\ 1,\ g_2(D) / g_1(D)\ ] \tag{2.20}$$

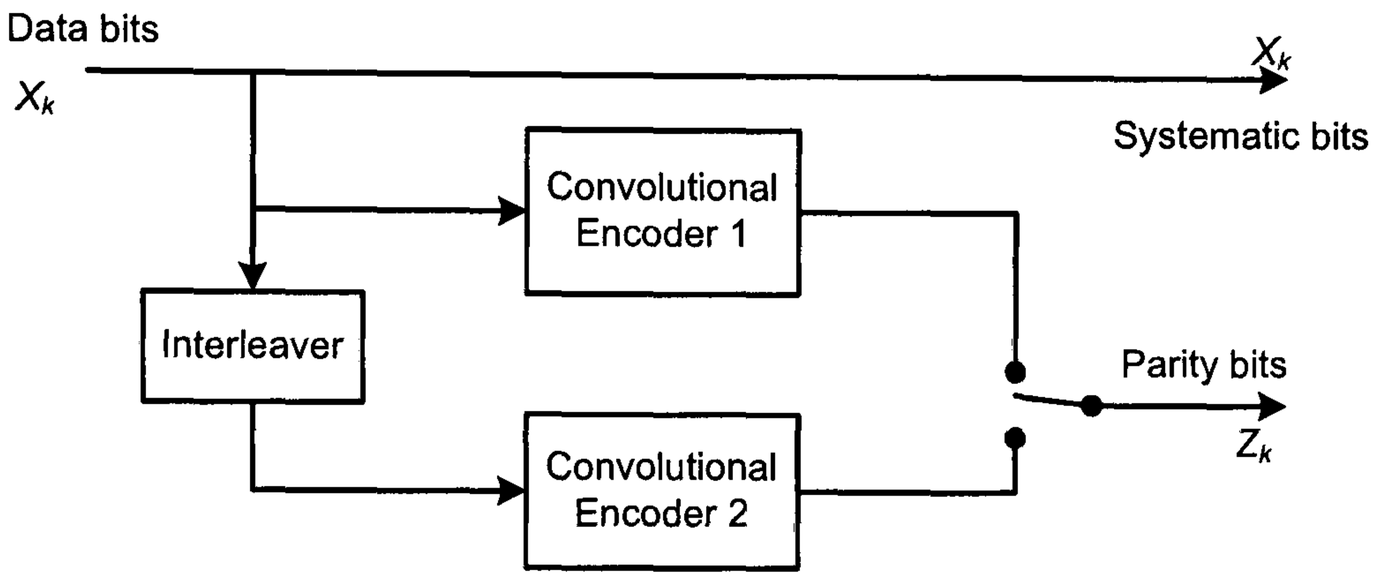


Figure 2-20: Turbo encoder [34]

Parity Bit Puncturer

Periodic puncturing of the parity bit stream has an effect on determining the compression efficiency of the DVC codec. For an input block of n bits, the two convolutional encoders at the turbo encoder generate $2n$ parity. The output of the DVC encoder is this parity bit stream for the input video sequence. Therefore, the parity bit sequence is subject to a periodic puncturing function, in order to achieve a compression. Selected parity bits are transmitted from the full stream in a periodic pattern.

The parity bit rate and pattern reflects the correlation noise in the side information with reference to the original Wyner-Ziv frame. Therefore, in common DVC frameworks, a feedback loop is utilised between the encoder and decoder. As a result, the parity puncturing for each block starts at a relatively small bit rate and the pattern is dynamically optimised by communication through the feedback channel.

2.4.3.3 Side Information Generation

The Wyner-Ziv decoder needs side information to reconstruct the current frame. Side information is statistically dependant on the original Wyner-Ziv frame sequence to be encoded. Because of the independent-encoding-joint-decoding concept of DVC, only previously decoded frames are available at the decoder for side information generation. Usually a sequence of reference frames, called key frames, is used in this process. The accuracy of the side information has a major effect on the parity bit rate: the more accurate side information, the lower number of bits needed for the decoding process.

The simplest method of side information generation is using the previous frame as side information. Performing average interpolation between the previous and next reference frames is another simple method. However, these simple techniques are often not effective enough because the temporal correlations are not sufficiently exploited. More complex techniques are therefore necessary to generate more accurate side information. Motion estimation and compensation algorithms have been an emerging solution among DVC researchers for accurate side information generation. In [35], motion compensated temporal filtering is used by Tagliasacchi et al. In [36], Natario et al. propose a motion field smoothing algorithm to generate side information. Ascenso et al. use forward and bidirectional motion estimation [17], [37] and propose a spatial smoothing algorithm by using a weighted median vector filter [17].

The motion interpolation technique used in DVC is quite different to those used in conventional video coding. A comparison of motion estimation process utilised in conventional video coding and DVC is given in Table 2.4:

Table 2.4: Comparison of motion estimation in conventional video coding and DVC [34]

Conventional Video Coding	DVC
Performed at encoder	Performed at decoder
Target frame is available for motion estimation	Target frame is not available for motion estimation
Actual motion field from reference frame to target frame is estimated.	Motion field is estimated across reference frames and interpolated or extrapolated.

Forward motion estimation is performed first between two reference frames. Following this, the calculated motion vectors are considered for bidirectional motion estimation followed by spatial

motion estimation. Finally, the predicted frame is obtained by using motion compensation. The following subsections explain each process in detail.

Forward Motion Estimation:

Forward motion estimation is the first step of frame interpolation, where the previous and next reference frames (X_{2i-1} and X_{2i+1}) are the only inputs. Input frames are first low-pass filtered to improve the reliability of the motion vector. A block matching algorithm is then used to estimate the motion between X_{2i-1} and X_{2i+1} as previously described in Section 2.2.2.2.

The main drawback of using forward motion estimation is the existence of overlapped and uncovered pixels in the interpolated frame because the motion vectors obtained by forward motion estimation do not necessarily intercept the interpolated frame at the centre of each non-overlapped block in the interpolated frame (see Figure 2-21). This problem is addressed by assigning motion vectors to each block of the interpolated frame in such a way that the previously obtained motion vectors serve as candidates. For each block, the motion vector that intercepts closer to its centre is selected (see Figure 2-22).

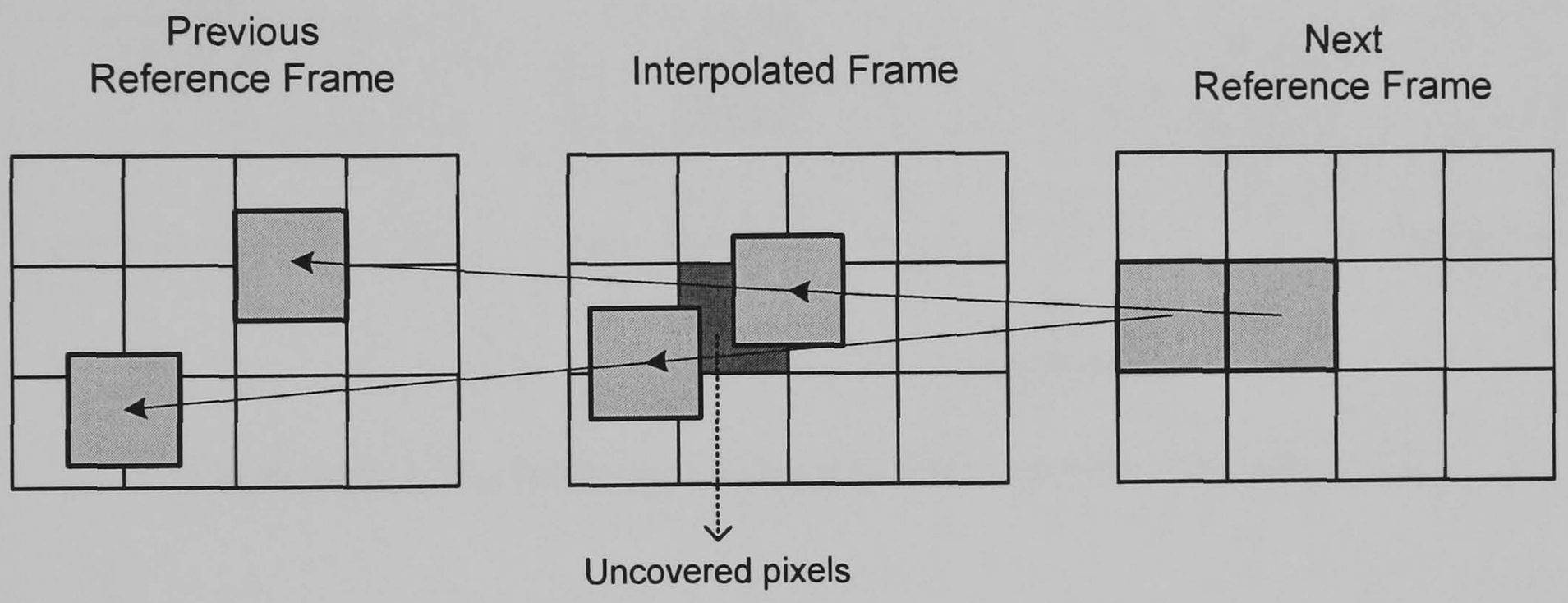


Figure 2-21: Uncovered pixels in forward motion estimation

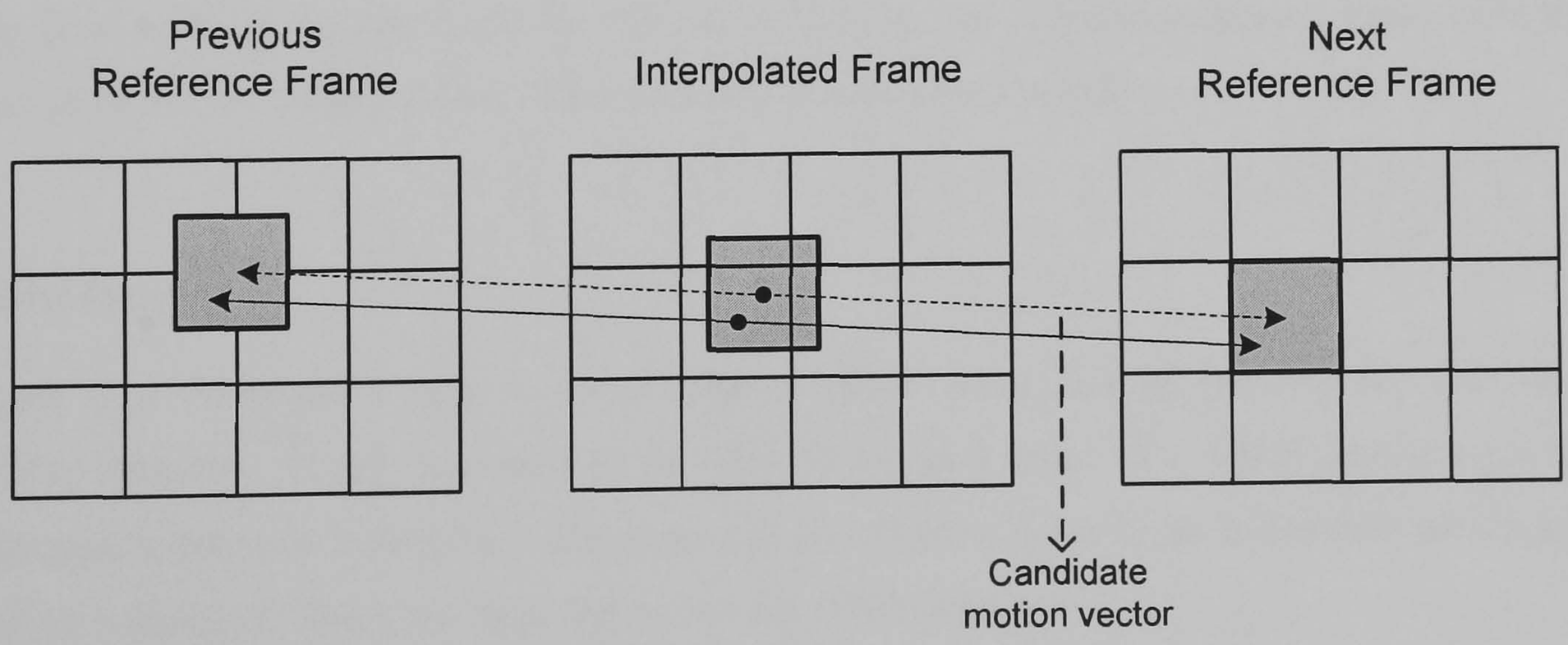


Figure 2-22: Selecting motion vector for a block of interpolated frame

Bidirectional Motion Estimation and Compensation:

The motion vectors obtained in the forward motion estimation process are refined using a bidirectional motion estimation algorithm. As previously discussed, unlike conventional coding techniques, the interpolated frame is not available in DVC. A linear motion is assumed between the previous and next reference frames, passing at the centre of the blocks in the interpolated frame as shown in Figure 2-23. A new search is performed around the candidate motion vector found using forward motion estimation and then the refined motion vectors are obtained. These refined motion vectors are filtered through the spatial motion smoothing algorithm using weighted median filters [38]. Finally, the motion vectors are used to fill the interpolated frame as in conventional video coding.

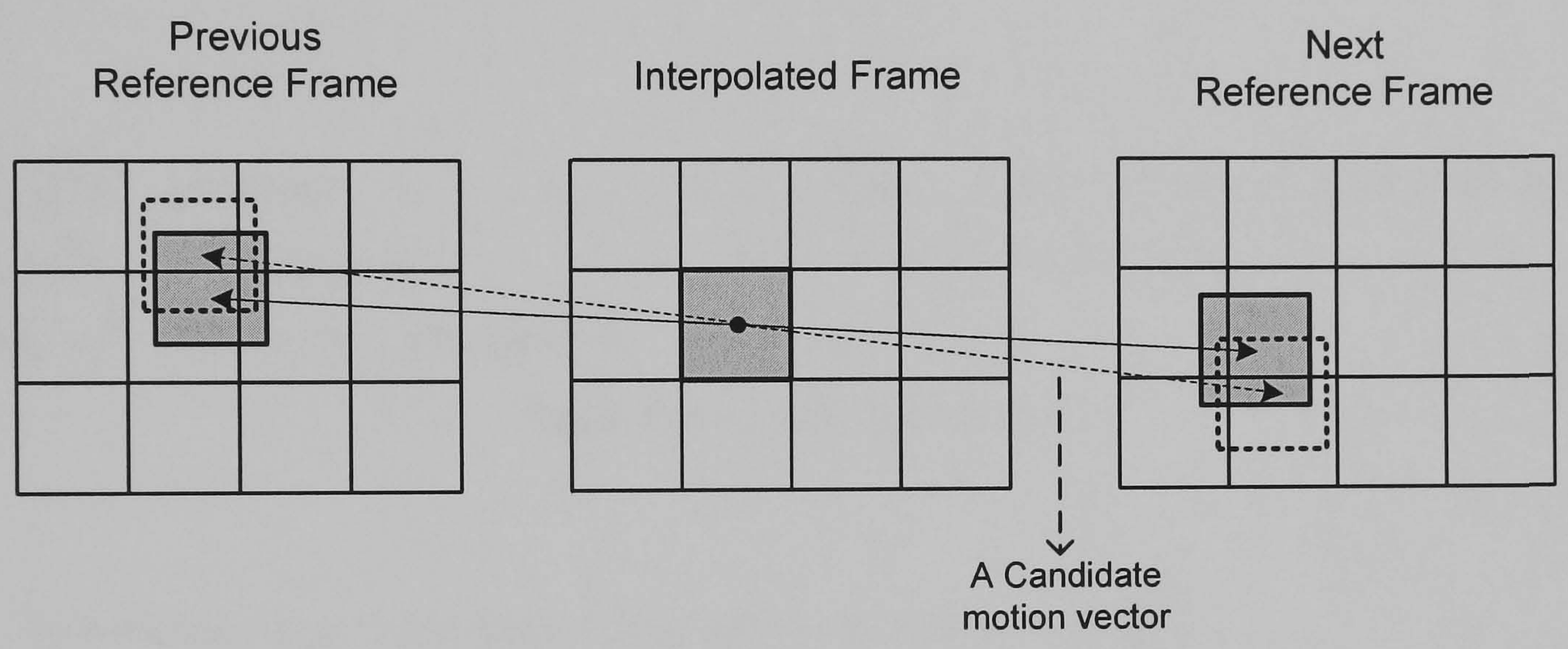


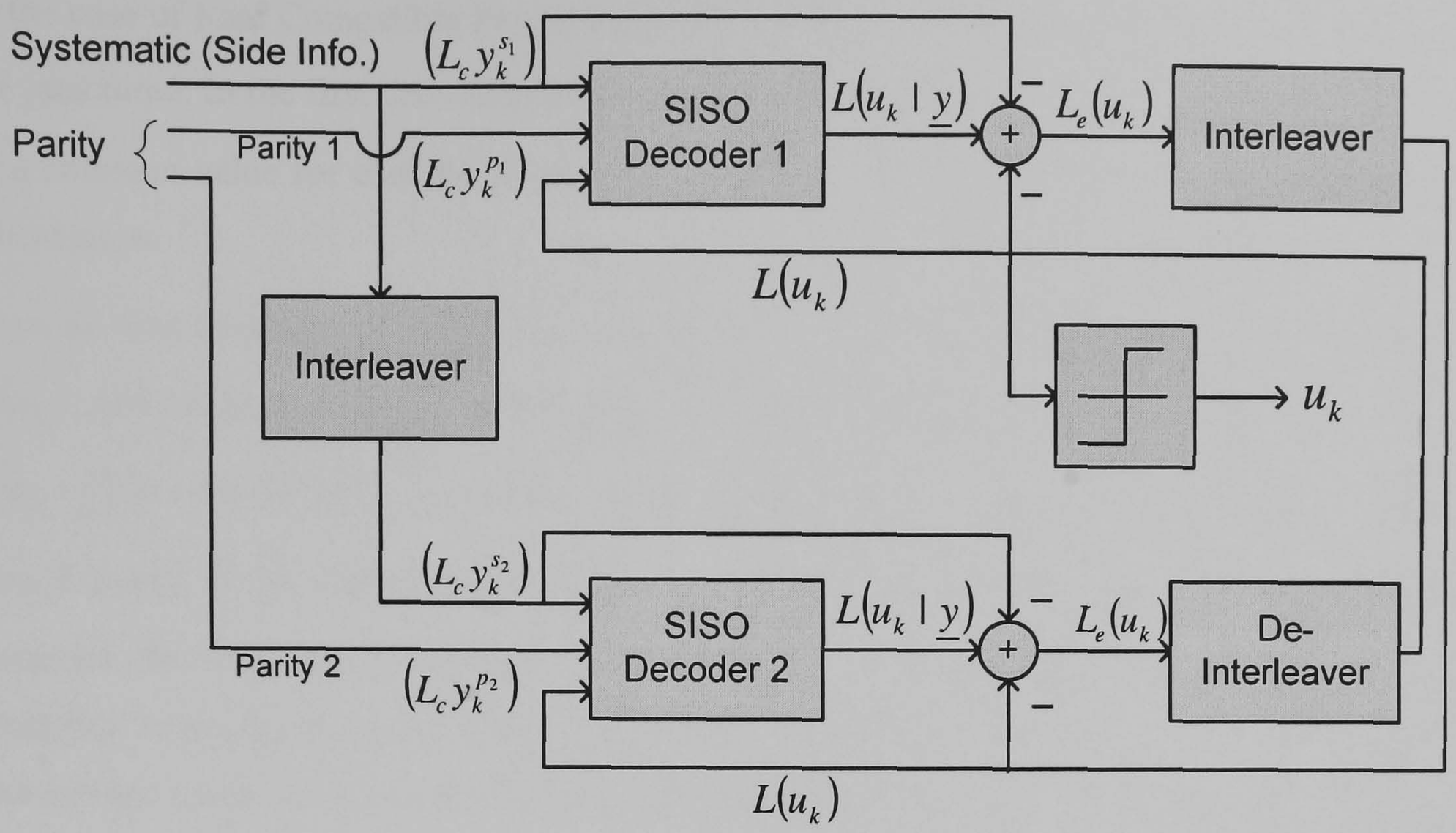
Figure 2-23: A candidate motion vector in bidirectional motion estimation

2.4.3.4 Slepian-Wolf Decoder

The Slepian-Wolf decoder includes the turbo decoder, the correlation noise estimation process and the error estimation process. These functions are discussed below.

Turbo Decoder

Figure 2-24 illustrates a turbo decoder implementation incorporating two soft-input soft-output (SISO) decoders. Turbo decoding is an iterative process since two SISO decoders exchange information between themselves. The term *soft information* refers to bit probability which allows making a decision. There are three inputs to each SISO decoder:



$(L_c y_k^p)$ - Parity bits

$(L_c y_k^s)$ - Systematic bits

$L(u_k \mid \underline{y})$ - *A-posteriori* information

$L(u_k)$ - Log likelihood ratio (LLR) of data bit u_k

$L_e(u_k)$ - Extrinsic LLR

Figure 2-24: Turbo decoder [34]

- Soft channel inputs consisting of the received parity stream $(L_c y_k^p)$

Where, y_k^p is the received punctured parity bit stream, L_c represents the *channel reliability* which is the noise modelling of the parity bit stream and L_c is dependant upon the channel Signal-to-Noise Ratio (SNR) and the fading amplitude in a typical communication scenario. Therefore L_c is derived through *channel estimation*. However, in DVC research an error free channel is usually assumed.

- Systematic bit stream $(L_c y_k^s)$

The systematic bit stream is calculated from the side information. Again, L_c represents the channel reliability. However, correlation noise modelling is utilised to estimate the L_c parameter since the *channel* is hypothetical with regard to the side information.

- The a-priori information $L(u_k)$

The *a-priori* information represents the reliability (a-posteriori) information generated by the other SISO decoder in the turbo decoder.

In the case of Rate Compatible Punctured Turbo (RCPT) codes that are used in DVC, parity bits are punctured. In the first decoding iteration, the Log Likelihood Ratio (LLR) $L(u_k)$ is initialised to a common value for each bit as there is no a-priori information available about the received information.

Once the first iteration of the SISO decoder is completed, the corresponding a-priori information $L(u_k)$ and input systematic information $(L_c y_k^s)$ are subtracted from the a-posteriori LLR $L(u_k \mid \underline{y})$ in order to derive the extrinsic LLR $L_e(u_k)$. This value is interleaved and used in the next iteration of the other SISO decoder as a-priori information $L(u_k)$. After a number of iterations, the LLR of the sequence usually converges to a value away from the hard decision boundary. Once the desired number of iterations is performed, the decoded soft information is sent through a hard decision filter to obtain the decoded bit stream.

Correlation Noise Modelling

The side information, an input to the Slepian-Wolf decoder, is usually modelled as a noisy version of the original Wyner-Ziv frame. Therefore, it is assumed that a hypothetical channel exists between the original Wyner-Ziv frame and the side information. Estimation of the statistical properties of this hypothetical channel applied to the systematic bit stream is critical for efficient performance of the turbo decoder. The correlation noise N between the original Wyner-Ziv frame X and the side information Y can be expressed as:

$$N = X - Y \tag{2.21}$$

This correlation noise is usually assumed to have a Laplacian distribution as in Equation (2.22), where α is the Laplacian distribution parameter [39]. Meyer et al. have reported that either a Gaussian or Laplacian noise model could be utilised for estimating the correlation noise [40]. In most DVC solutions, offline and ideal noise estimations have been utilised. The Laplacian parameter is estimated at sequence level (a common parameter for all frames), whereas a frame level or macro block level estimation can produce better results. Brites et al. have proposed an algorithm for dynamically estimating the noise parameters [39].

$$f(N) = \frac{\alpha}{2} e^{\alpha|N|} \tag{2.22}$$

Error Estimation

In feedback–based DVC decoders, after each decoding iteration, an error estimation is generated in order to decide whether more parity bits are needed for a preset output video quality. Therefore,

the decision criterion has an important effect on the bit rate. The most common solution is the *ideal error detection* method [1], where the decoded bit stream is compared to the original Wyner-Ziv bit stream. In this method, Bit Error Rate (BER) of the decoded bit stream is calculated with regard to the original Wyner-Ziv bit stream and more parity bits are requested until BER is below a preset threshold. This solution, although unrealistic as the original Wyner-Ziv frames are available at the decoder, is utilised by a majority of DVC researchers.

Tagliasacchi et al. have proposed a practical solution for terminating the bit rate request [41]. The Logarithm of the A Posteriori Probability (LAPP) ratio is monitored in SISO decoders; the higher the absolute value of LAPP ratio, the higher confidence in making the hard decision for a bit. Usually the average LAPP ratio (of all bits) increases after each iteration, thus the bit rate request is terminated when the average LAPP is more than a preset threshold.

2.4.3.5 Reconstruction

In DVC, the reconstruction module is utilised for inverse quantising the decoded and inverse bit plane extracted pixels while partly compensating for the quantisation noise. Side information, which is generated at the decoder, is used in addition to the decoded quantised pixels. Aaron et al. have proposed an initial solution that has been followed by a majority of subsequent DVC research due to its performance and simplicity of implementation. The reconstruction algorithm is described below:

Reconstruction for each pixel is described by one of the following three cases:

$$\hat{x}_i = \begin{cases} z^- , & y_i < z^- \\ y_i , & z^- < y_i < z^+ \\ z^+ , & y_i > z^+ \end{cases} \tag{2.23}$$

Where $\hat{x}_i$ is the reconstructed pixel, y_i is the corresponding side information pixel value, z^- and z^+ represent the lower and upper boundaries of the decoded quantised pixel value q'_i (as shown in Figure 2-25). When the side information is within the bin defined by the decoded quantised pixel, the output of the reconstruction function is equal to the side information. Otherwise, the output is clipped to either the lower boundary of upper boundary of that bin according to its value and the boundary values as illustrated in Figure 2-25. The reconstruction module can give better results if the side information and the decoded bit stream are highly correlated because the number of the pixels getting clipped will be lower. When the side information accuracy is low, the clipping effect will be reflected to the output by producing some annoying artefacts.

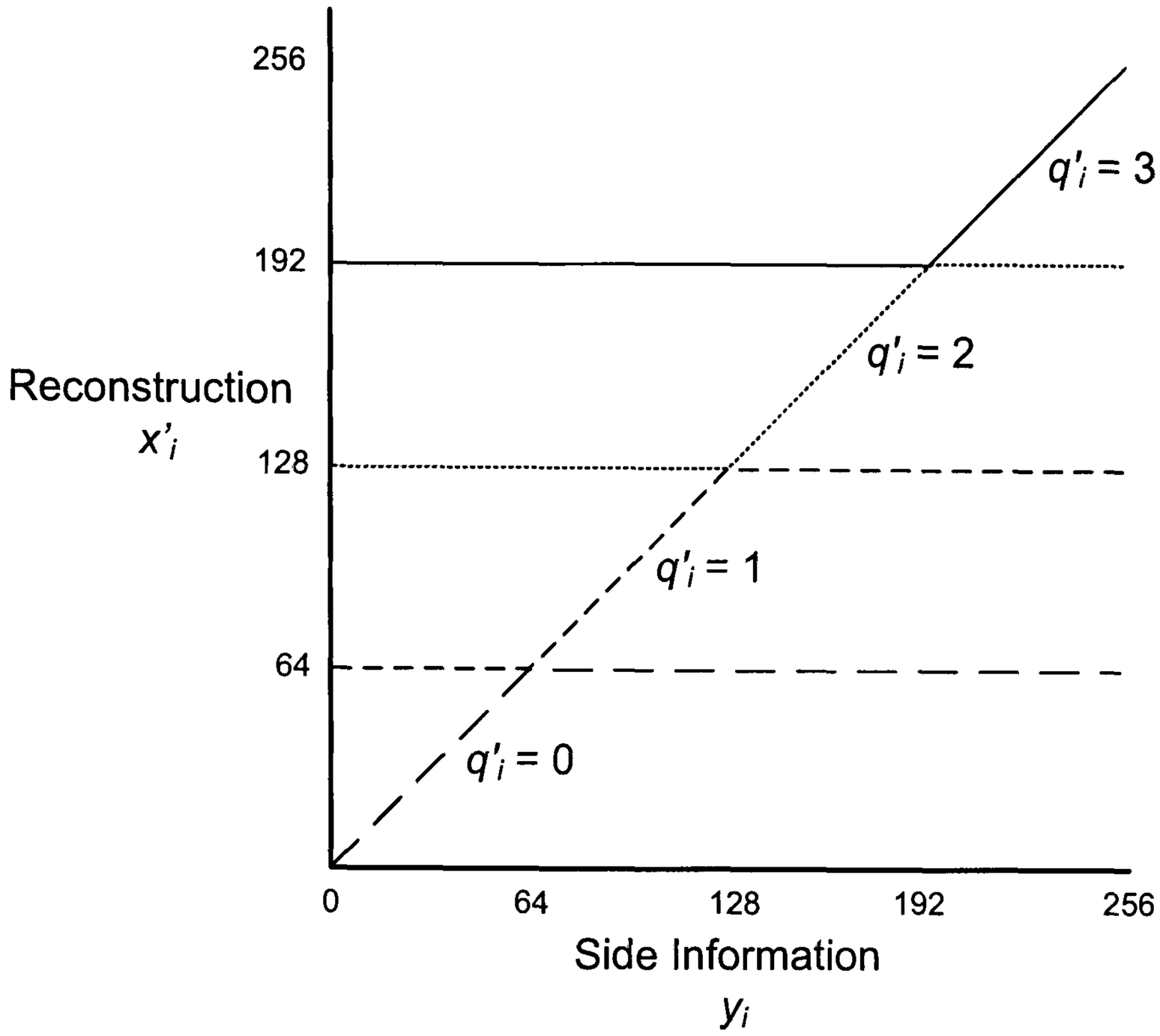


Figure 2-25: Reconstruction function for a 4-level uniform scalar quantiser [42]

2.4.4 Transform Domain DVC Codec Architecture

Transform coding is widely used in conventional video coding for efficient compression, hence its use in DVC. The Discrete Cosine Transform [43] is widely utilised in transform domain DVC implementations [15], [18] whilst the Wavelet transform has also been employed in some proposals [44]. The transform coefficients are then quantised in order to exploit the limitation of the HVS.

The transform domain DVC codec architecture is similar to the pixel domain architecture except for the DCT, inverse DCT (IDCT) and some differences in other blocks. Figure 2-26 illustrates a transform domain DVC codec using DCT. Usually a 4×4 block based DCT is used in transform domain DVC [15], [18] as defined by the H.264/AVC standard [11], which was described in Section 2.2.2.1. Once the DCT operation is performed at the encoder, each DCT coefficients band is independently encoded.

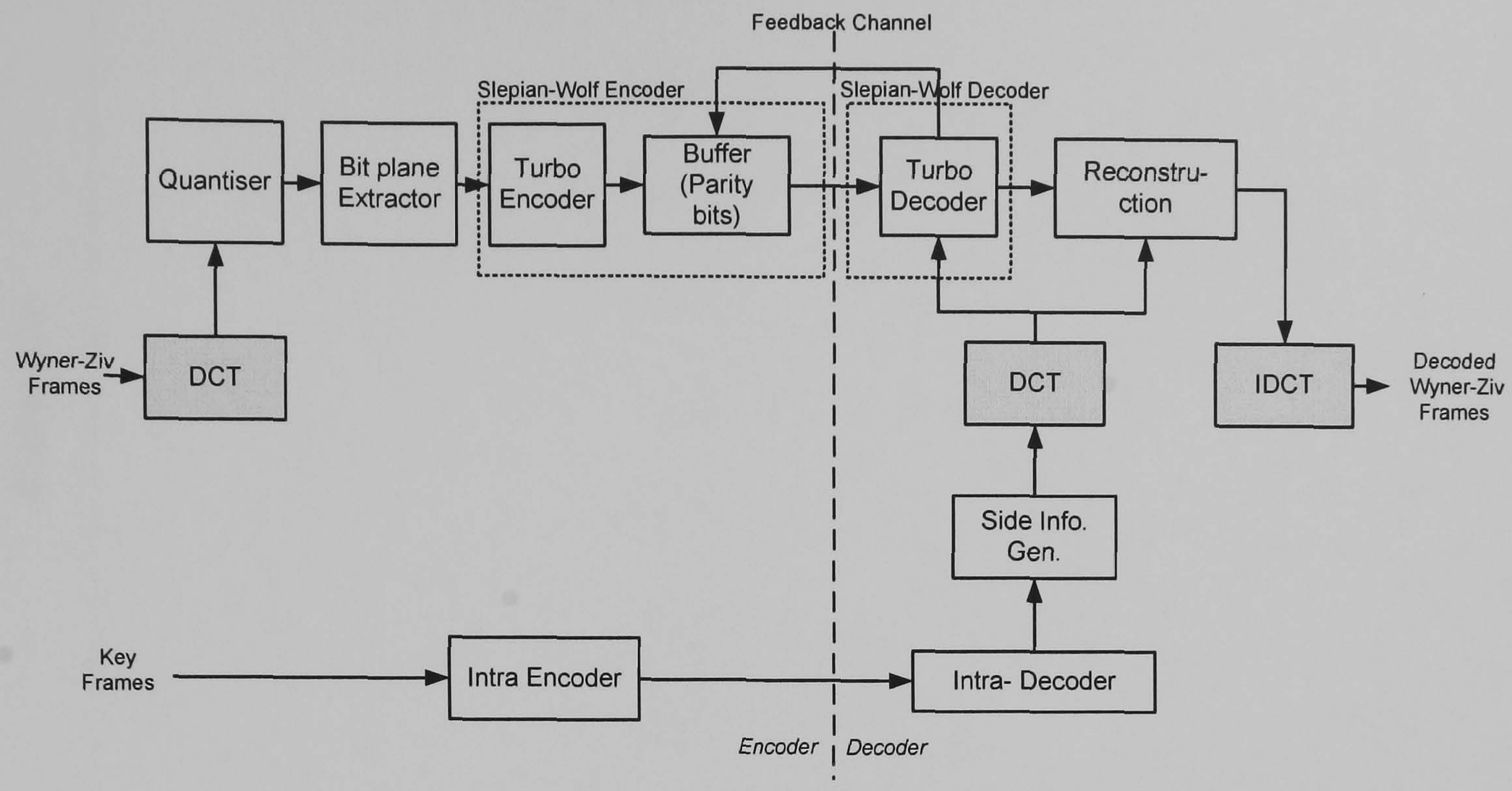


Figure 2-26: Transform domain DVC architecture

2.4.4.1 Transform Domain Quantisation

As in pixel domain DVC, a uniform scalar quantiser is utilised for transform domain DVC. However, in the pixel domain, only positive values are expected, whereas in the transform domain, DC coefficients are always positive and the remaining AC coefficients can be either positive or negative. Therefore, a different quantisation process is used for DC and AC bands in transform domain DVC.

For DC coefficient band quantisation, a similar algorithm to that used as in the pixel domain quantisation is utilised. For AC coefficient bands, the quantiser assumes both negative and positive values. Note that all AC coefficients are concentrated around zero [42]. The probability density function (pdf) for the lowest AC coefficient is illustrated in Figure 2-27. Given this density distribution, a uniform scalar quantiser with a symmetric quantisation interval around zero is utilised for quantising AC coefficient bands. The main reason for this is to increase the correlation between the quantised side information and the Wyner-Ziv bit stream by placing the highest probability at the centre of a quantisation bin. Figure 2-28 illustrates the quantisation for both DC and AC coefficient bands, where W is the quantisation step size and v is the quantisation input.

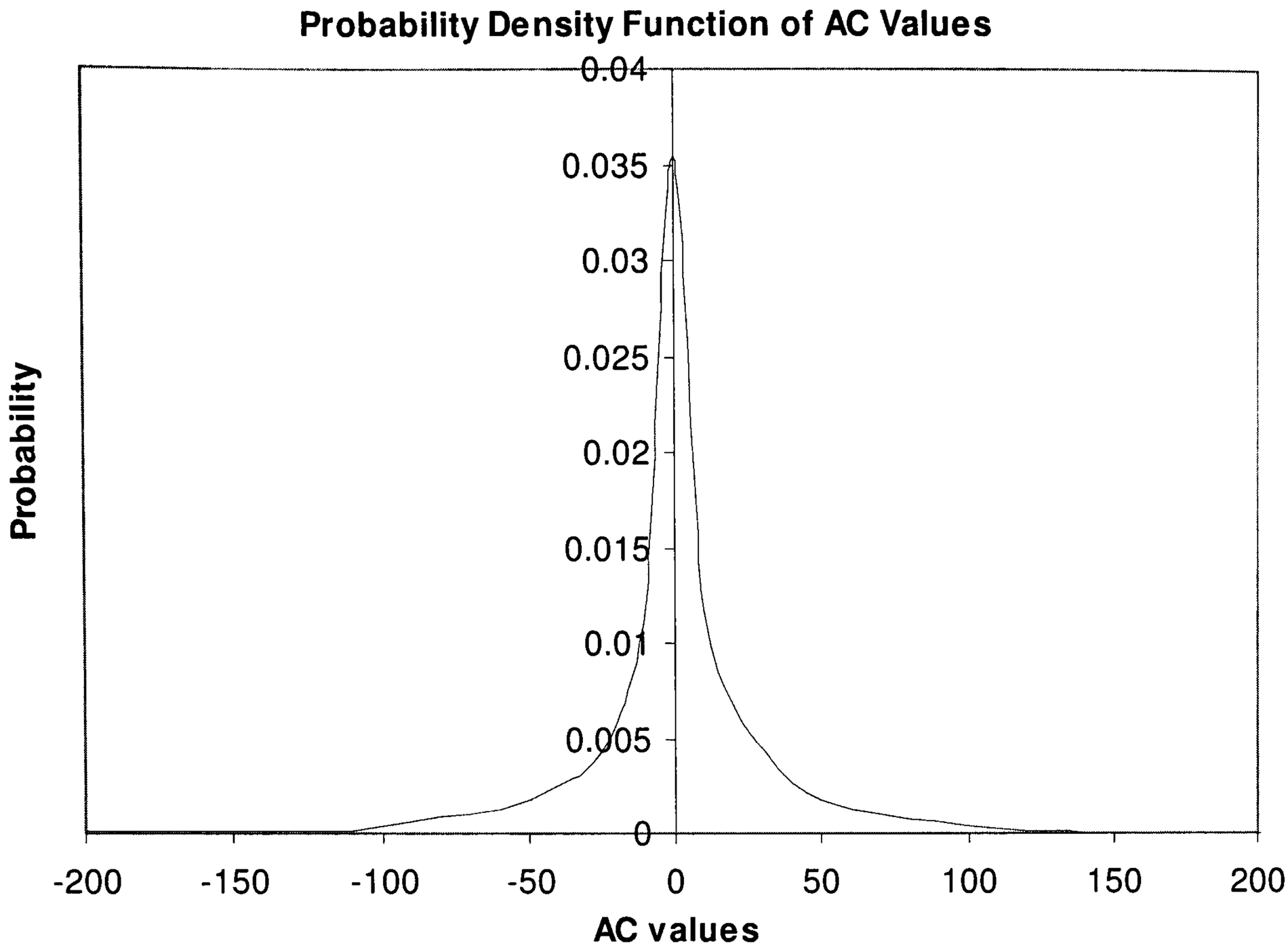
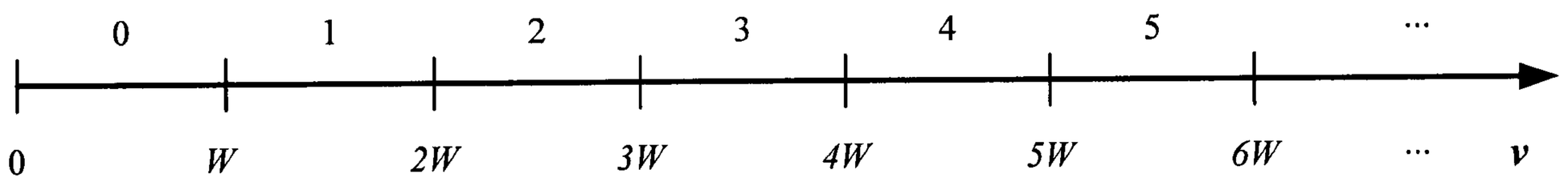
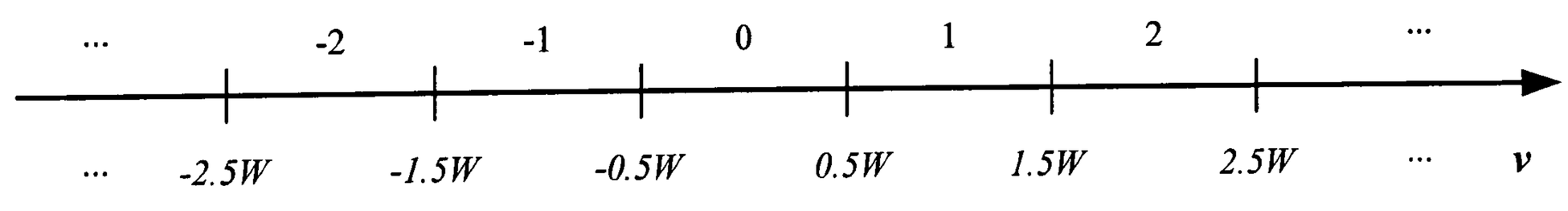


Figure 2-27: Probability density function of AC values (lowest spatial frequency) for *Foreman* QCIF sequence



(a)



(b)

Figure 2-28: Uniform scalar quantiser for (a) DC coefficients (b) AC coefficients [42]

The amplitude of AC coefficients tends to be lower for higher frequencies [42]. Additionally, the HVS is less sensitive to the higher spatial frequencies. Therefore, less quantisation levels are considered for higher frequency DCT coefficients. Eight different quantisation matrices—shown in Figure 2-29—are used for the quantisation of each DCT coefficient band; the values in the matrices show the number of levels [18]. The value 0 means no bits are transmitted to the decoder for the corresponding band. The number of quantisation levels is assumed to be known by both the encoder and decoder.

16	8	0	0
8	0	0	0
0	0	0	0
0	0	0	0

(Q1)

32	8	0	0
8	0	0	0
0	0	0	0
0	0	0	0

(Q2)

32	8	4	0
8	4	0	0
4	0	0	0
0	0	0	0

(Q3)

32	16	8	4
16	8	4	0
8	4	0	0
4	0	0	0

(Q4)

32	16	8	4
16	8	4	4
8	4	4	0
4	4	0	0

(Q5)

64	16	8	8
16	8	8	4
8	8	4	4
8	4	4	0

(Q6)

64	32	16	8
32	16	8	4
16	8	4	4
8	4	4	0

(Q7)

128	64	32	16
64	32	16	8
32	16	8	4
16	8	4	0

(Q8)

Figure 2-29: Quantisation matrices

Quantisation step sizes are calculated by dividing the dynamic range by the number of levels. The upper bound of the DC coefficient is defined as:

$$\sqrt{n^2} I_{\max} \tag{2.24}$$

Where, $n \times n$ is the DCT block size and $I_{\max}$ is the maximum pixel intensity. This is equal to 1024 for a 4×4 DCT size and 8-bit accuracy video.

For AC bands, the highest absolute value within each band is used as a range for quantisation. Thus, the step size can be calculated as follows:

$$W_k = \frac{2|V_k|_{\max}}{2^{M_k} - 1} \tag{2.25}$$

Where W_k is the step size for the AC coefficient bands, $k = \{2, \ldots, 16\}$, $|V_k|_{max}$ is the highest absolute value within each band and M_k is the quantisation level value for the same band. The decoder receives the dynamic range of each DCT coefficient [18] in order to synchronise the encoder and decoder.

Turbo decoding is performed independently for each DCT band and bit plane. Within each DCT band the most significant bit plane is decoded first as in pixel domain DVC. Reconstruction is then performed once all the coefficient bands have been decoded, in a similar manner to pixel domain DVC. Finally, the reconstructed DCT coefficients are inverse transformed and the output frame is obtained.

The DVC codec can work in both the pixel domain or transform domains depending on the application requirements. Pixel domain implementations allow lower complexity encoder designs compared to transform domain implementations because of the complexity of the transform coding operations. However, transform domain implementations result in a better rate distortion performance as some of the spatial correlations are exploited at the encoder.

2.5 Recent Developments in DVC

Side information generation is an area of high research interest for DVC because of its direct effect on the rate distortion performance. Recently a number of techniques have been published in this area. A mesh-based motion compensated interpolation approach has been proposed in [45]. A technique to improve performance by sending additional information to aid side information generation was proposed in [46]. A modified recursive search block matching approach was presented in [47]. An improved interpolation strategy with a multi-estimation mode framework was published in [48]. Tagliasaccchi et al. [49] perform a rate-distortion analysis of the DVC architecture, while focusing on the impact of the motion modelling on generating the side information. A non-stationary noise model for side information has been proposed in [50]. More recently, a novel frame interpolation method by using edge information was proposed in [51] to generate more accurate side information. Guo et al. have proposed a technique where a spatial splitting of the frames is performed to enable exploiting the spatial correlation at the decoder [52]. Zhang et al. have proposed a scheme using a full search for side-information to obtain the best

possible side information [53]. A rate-distortion analysis of motion-compensated interpolation at the decoder was presented in [54].

There have been several approaches using minimum Mean Square Error algorithm to improve the reconstruction algorithm in DVC [55], [56]. Weerakkody et al. have proposed an enhanced reconstruction algorithm where the side information is more effective when the error rate is higher [57].

Key frames are often coded using conventional video coding techniques (lossy coding). However, key frames can be assumed to be losslessly available at the decoder in order to concentrate on Wyner-Ziv coding, although this is not a realistic option. JPEG 2000 [58] and H.264/AVC intra-coding [59] are often utilised in lossy key frame coding scenarios. However, it is shown in [60] that coding key frames using H.264/AVC intra-coding performs better compared to JPEG 2000.

In different implementations of DVC, a 3D scene-modelling scheme is proposed in [61]. Artigas et al. have proposed side information generation through a fusion approach for multi-view DVC scenarios [62]. Error resilience of DVC has been discussed in [2], [63] and [64].

Even though a lot of work has been done, still considerable research input is needed to improve the rate distortion performance of the DVC codec.

2.6 Conclusion

In this chapter a detailed review of the literature has been presented. First, the video coding basics are presented starting from video signal to the compression techniques. Conventional video coding is then introduced by presenting the encoder and decoder architectures. Afterwards, the theoretical bases of DVC, Distributed Source Coding principles, are explained. Encoder-decoder complexity balance of DVC is compared to that of conventional video coding. Pixel and transform domain DVC codec architectures and components are discussed in detail. Finally some recent publications are cited.

Chapter 3

3 Side Information Refinement

One of the major challenges in distributed video coding is the prediction of the Wyner-Ziv frame at the decoder. Because, the key to good decoding performance in distributed video coding systems lies in the efficient prediction of frames using side information. The more accurate the side information is the fewer bits need to be transmitted for the decoding process. Most DVC frameworks perform this process without considering progressive nature of the synthesis of the final frame. Therefore, the accuracy of the prediction drastically falls for high motion sequences and higher group of picture sizes because of the lack of information about the frame to be decoded. Several iterative side information refinement techniques have been reported in the literature, which are discussed in the following section (Section 3.1). However, there is lots of room available to exploit source correlations through iterative side information refinement.

Observing the fact that incorrectly predicted areas of current frame can be detected at different levels of final frame synthesis, several algorithms for refinement of side information synthesis based on an additional stage of motion estimation are proposed in this chapter. The refinement is performed on correctly decoded partial frames and is used to significantly improve the motion prediction of the final resolution frames which leads to enhanced performance of the overall system. An iterative refinement is performed at the decoder, first with the help of decoded DC frame and then considering partially decoded frame using previously refined side information. By iteratively refining the side information, a significant improvement has been achieved in the rate distortion performance.

In this chapter, also a novel modified framework for the DVC codec is proposed, considering intra-coding of the DC Wyner-Ziv frame and improved side information generation with the help of the decoded DC frame. A significant improvement has been achieved in the rate distortion performance by the proposed side information generation technique with the help of intra-coded DC frames.

The rest of this chapter is organised as follows: related work is given is Section 3.1, DC frame and oversampled DC frame is explained in Section 3.2, proposed DC refinement technique and iterative refinement technique are presented in Sections 3.3 and 3.4, the proposed intra-coded DC frame algorithm is given in Section 3.5 and finally Section 3.6 concludes this chapter.

3.1 Related Work

In 2005, Artigas et al. proposed a DVC codec which makes use of the decoded information to improve the side information [1]. In this technique, first standard motion compensated temporal interpolation is used to get the initial Side Information. Once the WZ frame is decoded, this first output is called "partially decoded picture". Then, conventional motion estimation is performed as the current frame is available (partially decoded picture). The new side information is fed into the DVC decoder again this loop can be repeated several times. This technique targets to improve the quality without changing the bit rate as the refinement is performed after the decoder receives all the necessary parity bits from the encoder.

In 2006, Adikari et al. proposed a refinement technique using multiple side information streams in [66]. Side information is generated by sequential motion compensation using both luminance and chrominance information together. Later this work was extended using multiple side information streams [67] and 3D motion refinement [68], Previously decoded bit planes (higher significance) are iteratively used to perform the motion estimation (in luminance and chrominance spaces) so that more improved side information for subsequent bit planes is produced. The new motion field derived in this step is used for motion compensating the same reference frame.

In 2007, Weerakkody et al. proposed another side information refinement algorithm exploiting the spatial correlations in the partially decoded current Wyner-Ziv frame in [69]. In this algorithm, first the frame is interleaved to suppress the adverse effects of burst errors. Then, the error estimation results are compared against a pre-determined error threshold to flag erroneous blocks. The frame is de-interleaved to scatter flagged bit bursts. Flagged bits in the frame are re-predicted using a prediction algorithm. The resultant frame is then interleaved again and several iterations of error estimation, flagging and prediction steps are performed.

In 2008, we proposed a novel transform domain side refinement algorithm based on DC refinement [70] and later in 2009, this work was extended for AC coefficients [71] which are presented and discussed in details in the following sections.

In 2009, S. Ye et al. proposed a side information refinement exploiting spatial and temporal correlations in [72]. They considered partially decoded frame to improve the side information. Also, motion vector refinement and smoothing, optimal compensation mode selection, and a new matching criterion for motion estimation algorithms were presented in this work. Motion estimation is based on a matching criterion using both mean absolute difference and boundary absolute difference.

3.2 DC Frame and Oversampled DC Frame

In DVC, predicting the side information (SI) accurately lies in efficient use of the available information at the decoder. Exploiting the additional available information during the decoding process is the main idea of this chapter. This is achieved by iteratively refining the side information as further DCT coefficients (of the whole frame) are decoded. The core of the proposed techniques in this chapter is the side information refinement when the decoded DC frame is available as well as the previous and next reference frames. And the most critical process for the side information refinement technique is the motion estimation in DC domain.

Side information refinement algorithm is first applied after the DC coefficients are decoded because some information about the current WZ frame becomes available at the decoder. The DC coefficients are grouped together and this group is called *DC frame* (D) (Equation (3.1)) which has a size of ($m/b \times n/b$), where ($b \times b$) is the DCT size and ($m \times n$) is the original frame size. Key frames need to be in DC domain in order to estimate the motion in DC domain. However, when the key frames are DC transformed, the DC key frame does not contain as much details as the original frame. And also the motion field resolution is lower as these motion vectors will be used in compensating the original size frame. The most effective way to solve this problem is using Oversampled DC frame ($\tilde{D}$). Oversampled DC frame is obtained by DC transforming all possible $b \times b$ blocks (sliding $b \times b$ window pixel by pixel) of the original frame (Equation (3.2)) and as a results oversampled DC frame size is (m-b+1) $\times$ (n-b+1) (see Figure 3-1 and Figure 3-2).

$$D(i,j) = \sum_{n_1=0}^{b-1} \sum_{n_2=0}^{b-1} X(bi + n_1, bj + n_2) \tag{3.1}$$

$$\tilde{D}(i,j) = \sum_{n_1=0}^{b-1} \sum_{n_2=0}^{b-1} X(i + n_1, j + n_2) \tag{3.2}$$

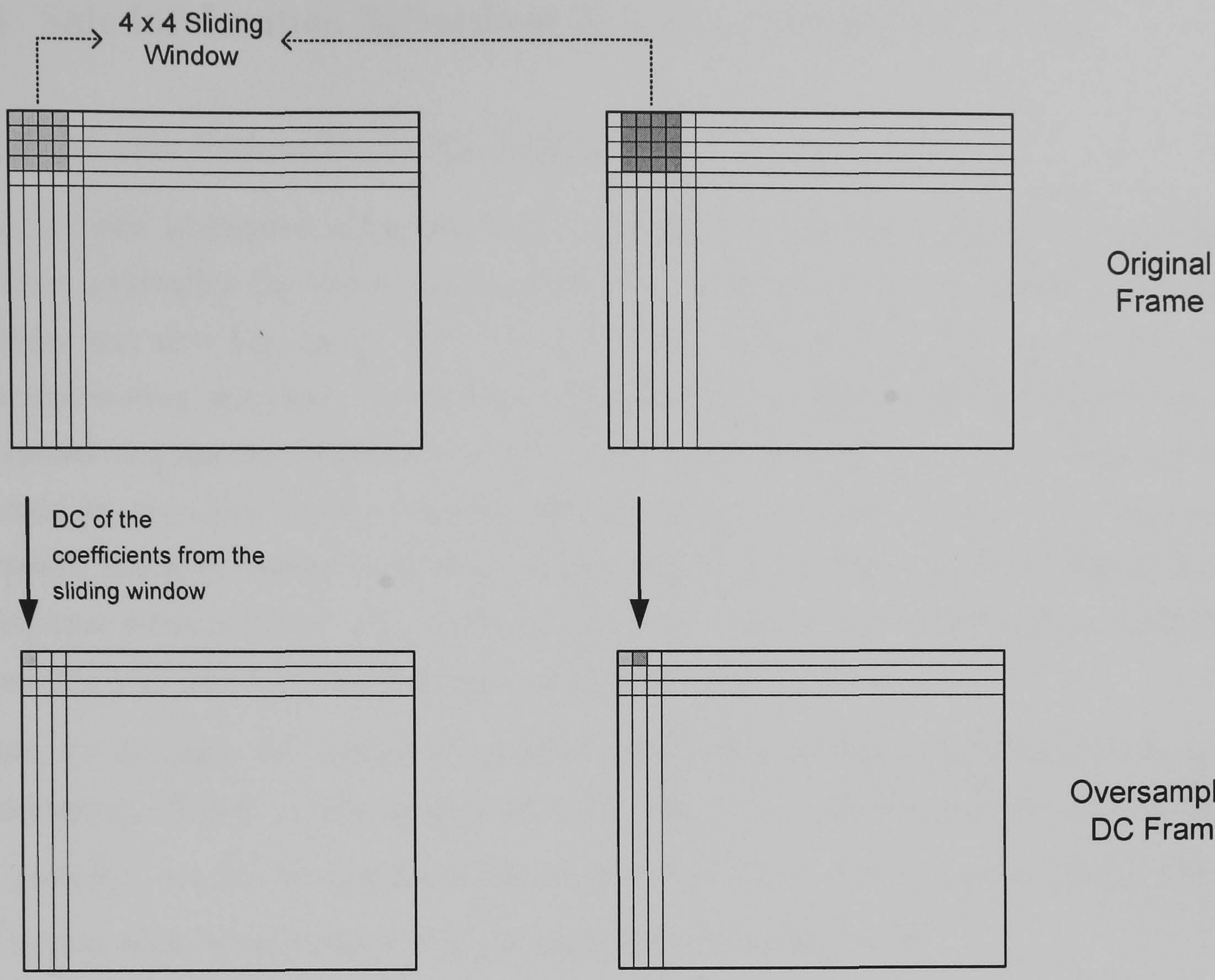


Figure 3-1: Forming oversampled DC frame

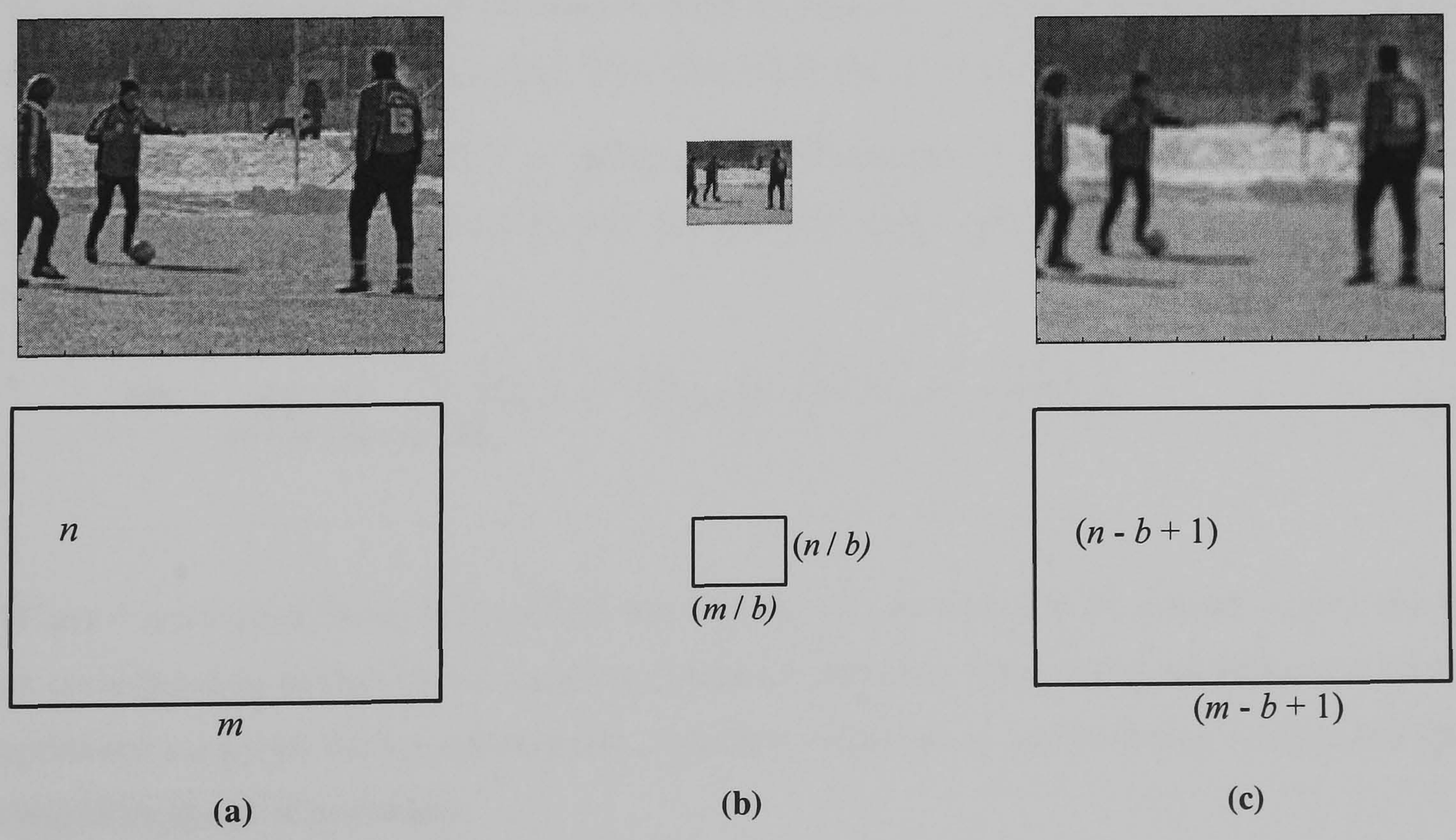


Figure 3-2: Frames and their dimensions for (a) Original frame (b) DC frame and (c) Oversampled DC frame (b and c normalised)

3.3 Side Information Refinement Using the Decoded DC Frame

3.3.1 Motion Estimation in DC Domain

Proposed side information refinement codec architecture is depicted in Figure 3-3. As in the previous approaches the initial side information is obtained by motion interpolation using previous and next key frames [18]. The initial side information is generated without any information about that frame. The initial side information is DCT transformed and the decoding is performed bit-plane by bit-plane, from DC to AC coefficients. First, the DC coefficients are decoded by the turbo decoder using the parities received from the encoder. In the proposed approach, instead of directly proceeding with decoding the AC coefficients, the conventional side information frame is refined using additional motion information extracted from the decoded DC frame (D_k) in an additional refining motion estimation stage.

Figure 3-4 illustrates the refinement algorithm, where ME is motion estimation, MO is motion optimisation, OS DC is oversampled DC, X'_{k-1} and X'_{k+1} are the reference key frames, $\tilde{D}_{k-1}$ and $\tilde{D}_{k+1}$ are the oversampled DC frames of the key frames, **mv** is the motion vector, **mm** is the motion mode for each motion block, S' is the refined side information.

In the proposed scheme, the refining stage of motion estimation is performed in DC domain because of the higher quality (successfully decoded) available at the decoder. The improved prediction for the higher resolution is obtained via motion estimation between the decoded DC WZ frame D_k and transformed key frames. Reliable matches to the decoded DC frame blocks are achieved by considering oversampled DC frames of the key frames.

The motion search (Figure 3-5) is performed for all blocks of the DC frame, separately for backward and forward directions ($p = \pm 1$ for GOP=2) using sum of absolute difference (SAD) metric for block-matching:

$$\mathbf{mv} = \underset{\mathbf{mv}=(mv_x, mv_x)}{\arg\min} \sum_{(i,j)\in \mathbf{B}} \left| D_k(i,j) - \tilde{D}_{k+p}(b\cdot i + mv_x, b\cdot j + mv_y) \right| \tag{3.3}$$

Where k is a current frame index, **B** is a block in D_k , (i, j) is a pixel in **B,** and $\mathbf{mv} = (mv_x, mv_y)$ is its corresponding motion vector for either prediction direction. The motion information is further optimised using the motion optimisation algorithm explained in the following section, and then final SI synthesis is performed.

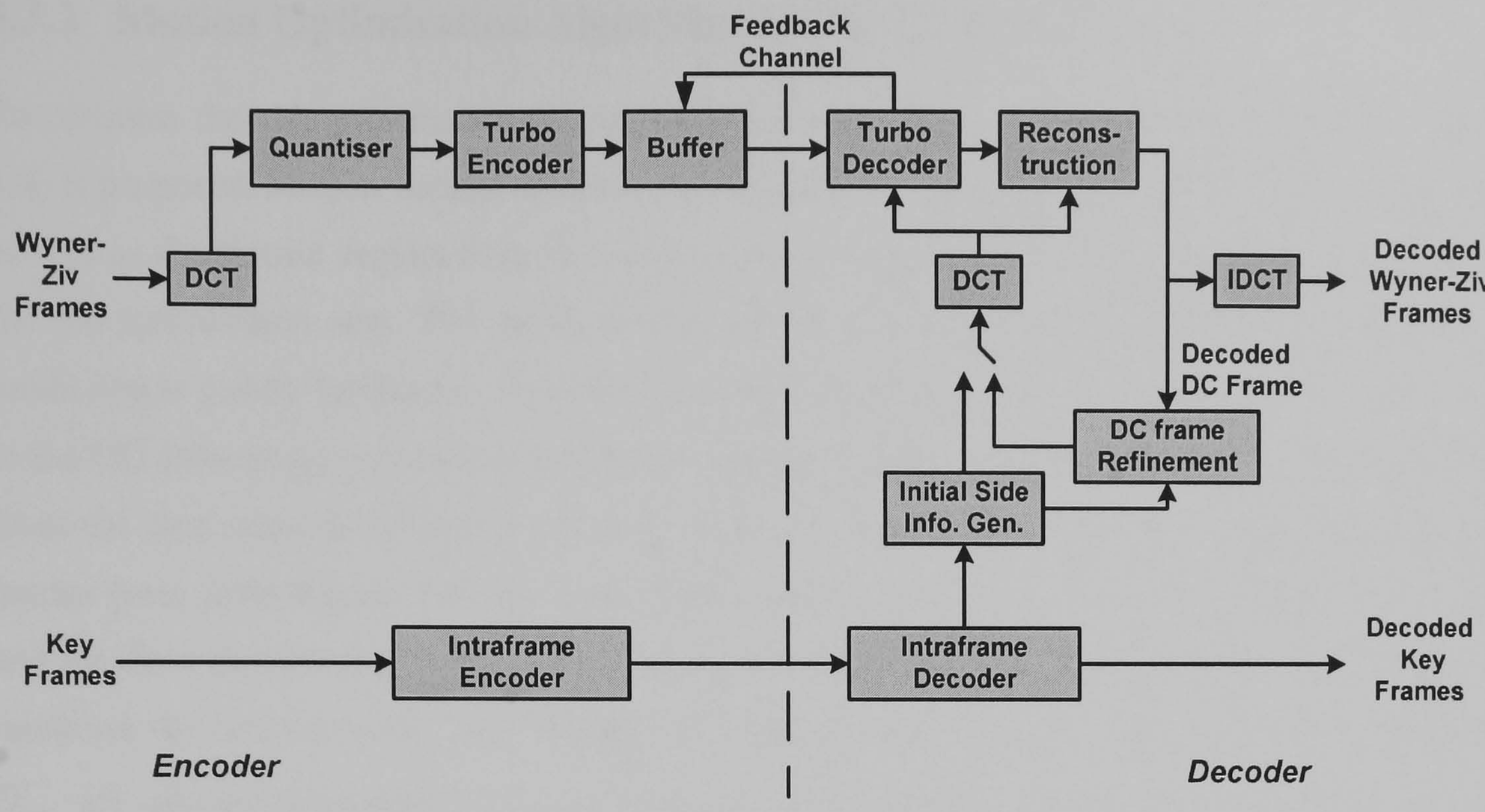


Figure 3-3: Proposed SI refinement codec architecture

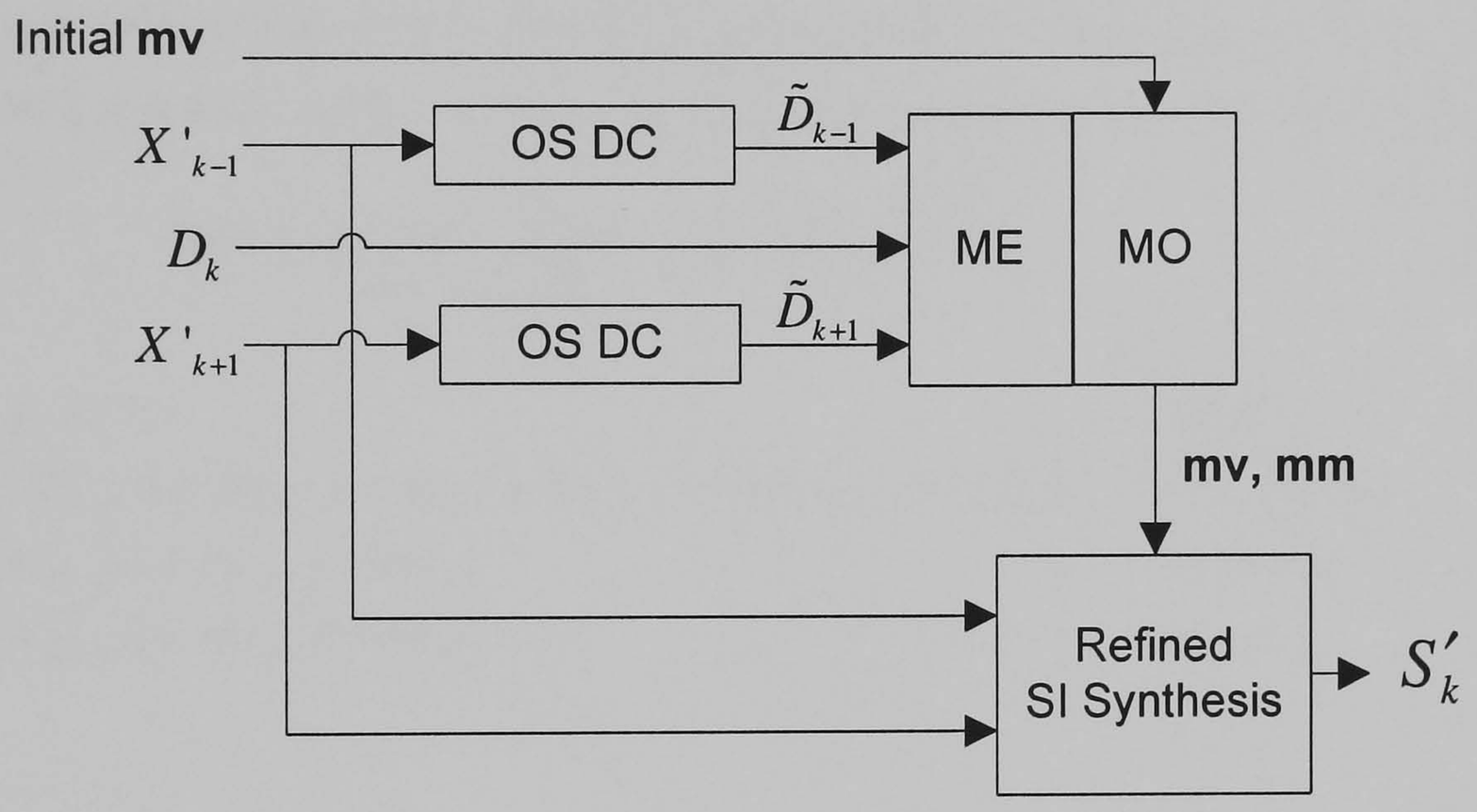


Figure 3-4: Proposed DC frame refinement algorithm

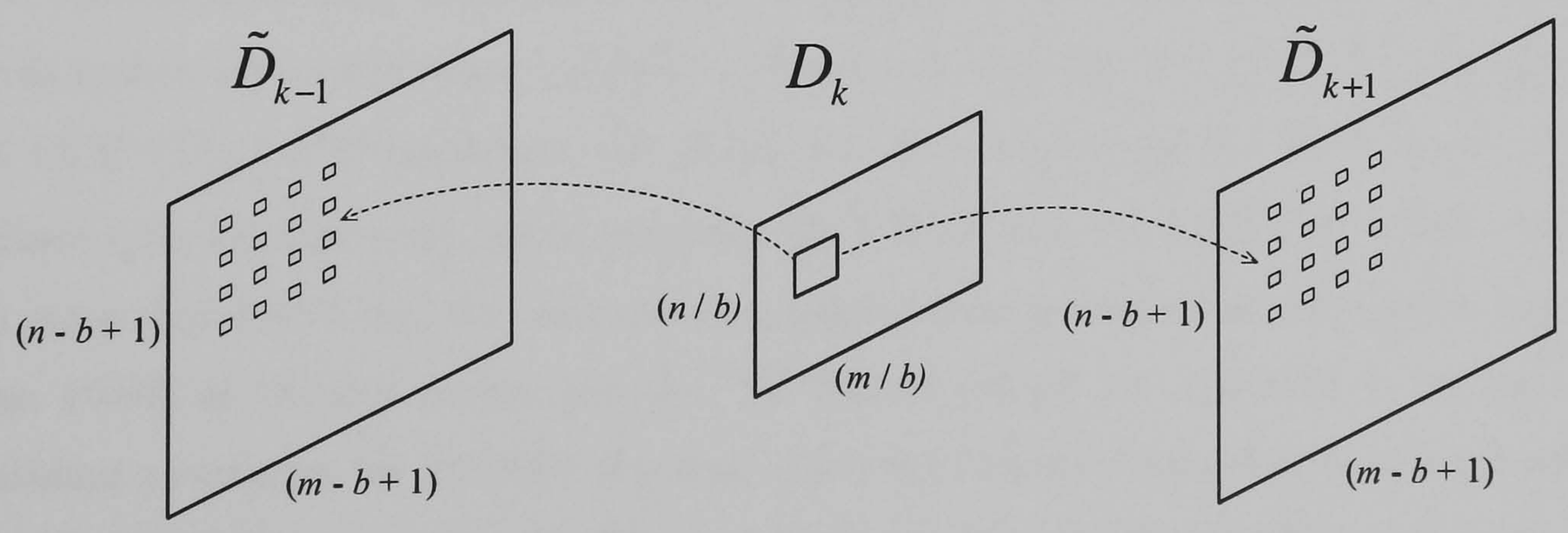


Figure 3-5: Motion estimation in DC domain for side information refinement

3.3.2 Motion Optimisation Algorithm for SI Refinement

To optimise the side information quality a novel multi-mode selection technique (MO in Figure 3-4) is proposed. Motion vectors obtained for forward and backward prediction in the DC domain, as well as the motion vectors from the initial motion interpolation are pre-selected, i.e. considered for the optimisation step. The initial motion vectors are considered because their bidirectional prediction is jointly optimised, while the independently predicted motion vectors from the search in the DC domain are particularly useful for identifying and correcting incorrectly predicted areas from the first stage of prediction. At this stage, for each block of the final frame three motion modes (**mm** from Figure 3-4) are considered - forward, backward and bidirectional predictions and for these modes all pre-selected motion vectors are evaluated in DC domain. The modes that minimise the SAD between the decoded DC frame D_k and predicted frames by different modes $D'_{\mathbf{mm}}$ are selected (Equation (3.4)). The selected mode and corresponding motion vectors are used for the final synthesis of refined side information (S'_k from Figure 3-4) at full resolution (Equation (3.5)). Refined side information is then DCT transformed in order to support decoding of the remaining DCT and then finally the frame is reconstructed using refined side information.

$$\mathbf{mm} = \arg\min_{\mathbf{mm}} \sum_{(i,j)\in\mathbf{B}} \left| D_k(i,j) - D'_{\mathbf{mm}}(i,j) \right| \tag{3.4}$$

$$S'_k(i,j) = \begin{cases} S_k(i,j) & \text{, initial} \\ \left(X'_{k-p}(i+mv_x, j+mv_x) + X'_{k+p}(i+mv_x, j+mv_x) \right) / 2 & \text{, bidirectional} \\ X'_{k-p}(i+mv_x, j+mv_x) & \text{, backward} \\ X'_{k+p}(i+mv_x, j+mv_x) & \text{, forward} \end{cases} \tag{3.5}$$

3.3.3 Experimental Results

The proposed method has been integrated in VISNET II codec [6]. In the experiments *Foreman*, *Soccer*, *Hall-Monitor* and *Coastguard* QCIF sequences are considered because of containing different motion levels and being the most common test sequences for DVC [5]. The sequences are of QCIF (176×144) resolution and 15 fps. Results are obtained for GOP size of 2. In all experiments the settings were: luminance only, DCT block size 4×4. The key frames are intra-coded using H.264/AVC and the corresponding quantisation parameters are selected to match the average PSNR of the key frames and the WZ frames (please see appendix C for key frame quantisation parameters for VISNET II codec). Different rate distortion (RD) points are obtained using eight quantisation matrices in Figure 2-29 (page 42), which are widely accepted in the transform domain DVC codecs [18].

The results are given in Figure 3-6. For higher motion containing sequences, *Foreman* and *Soccer*, the DC refinement technique outperforms the current solution [6] by up to 1 dB. However, for lower motion sequences, *Coastguard* and *Hall-Monitor*, the improvement by the DC refinement technique is smaller. This is because of the inaccuracy of the side information generated by motion interpolation algorithm for the high motion sequences.

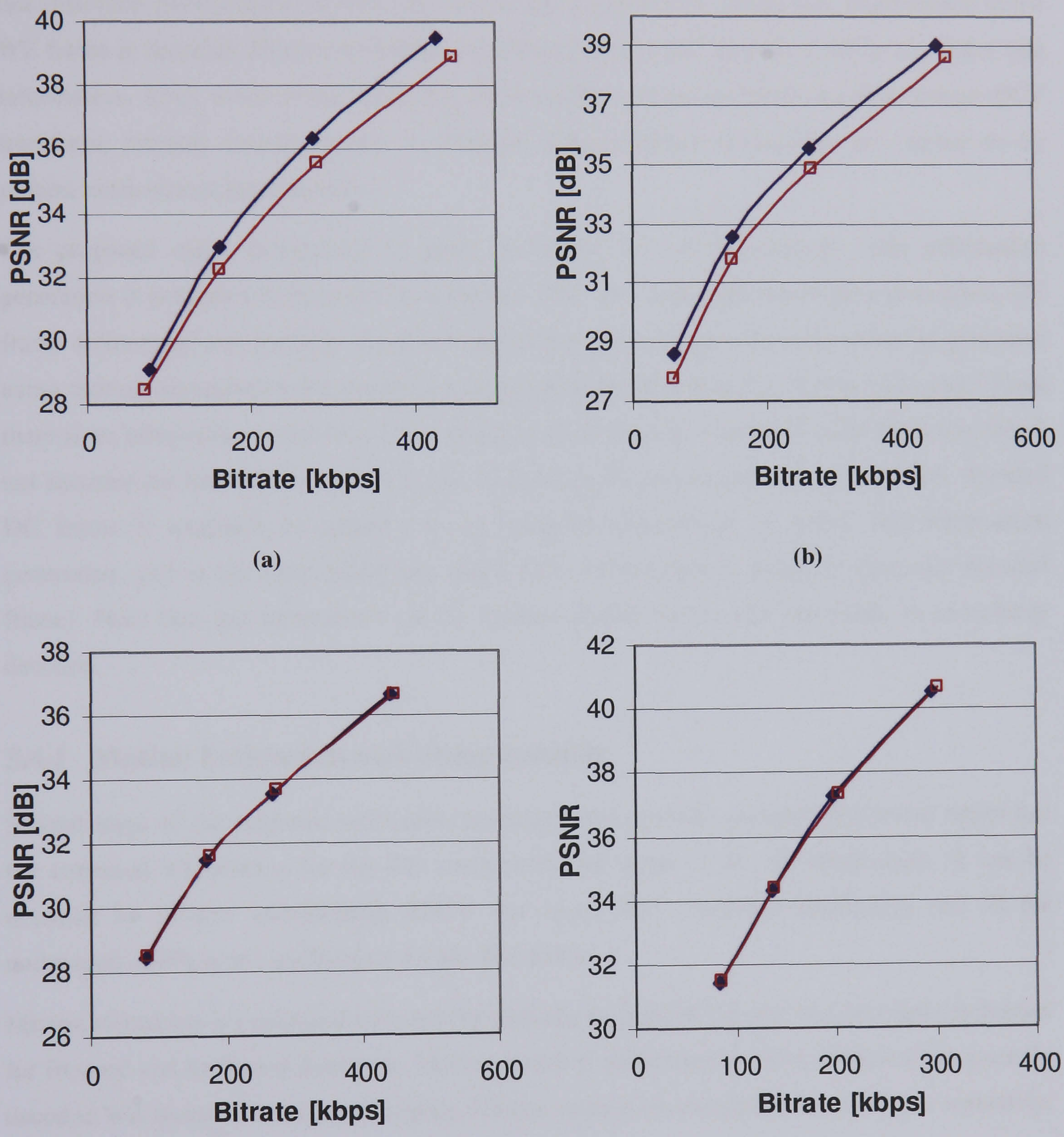


Figure 3-6: RD comparison for (a) *Foreman* and (b) *Soccer* (c) *Coastguard* and (d) *Hall-monitor* sequences

3.4 Iterative Side Information Refinement

In the previous section a side information refinement technique using DC refinement is presented. This method can be further improved by considering other decoded information i.e. AC coefficients iteratively. Therefore, the technique presented in this section aims to decrease the AC coefficients bit rate, while keeping the DC coefficients quality as it is and increasing the overall rate distortion performance. Iterative refinement starts after the DC image (DC coefficients) of the WZ frame is decoded. Motion estimation is performed in the DC domain to refine the initial side information. Then, some of the remaining DCT coefficients are decoded and after inverse DCT transform, partially decoded frame is obtained. Other iterative refinements are carried on by motion estimation in pixel domain.

The proposed codec architecture is given in Figure 3-7. At the decoder, side information generation is presented in three different blocks. They are initial side information generation, DC frame refinement and partially decoded frame refinement. Initial side information is generated using motion interpolation between the decoded previous and next key frames as in [18]. Since there is no information about the current frame, a linear motion is assumed [18]. However, this is not accurate for fast and non-linear motion sequences. In the second refinement block, decoded DC frame is available in addition to all available information to initial side information generation, and in the third refinement block more information is available (partially decoded frame). Note that last refinement can be applied several times until the frame is completely decoded.

3.4.1 Motion Estimation and Compensation

Second stage of the proposed refinement technique uses partially decoded WZ frame which has the corrected information for the DC coefficients and some of the AC coefficients. It can be obtained by inverse transforming (IDCT) the latest DCT (decoded coefficients and SI for undecoded coefficients) coefficients for the WZ frame.

Motion estimation is performed between the partially decoded WZ frame and decoded key frames for forward and backward directions. Motion search is performed at full resolution since partially decoded WZ frame is in the pixel domain. Motion vectors are calculated for all blocks separately, using:

$$\mathbf{mv} = \underset{\mathbf{mv}=(mv_x,mv_x)}{\arg\min} \sum_{(i,j)\in\mathbf{B}} \left| Y_k(i,j) - X'_{k+p}(i+mv_x, j+mv_y) \right| \tag{3.6}$$

Where Y_k is the partially decoded WZ frame, X'_{k+p} is the decoded key frame, B is a block in Y_k and **mv** is its corresponding motion vector for either prediction direction.

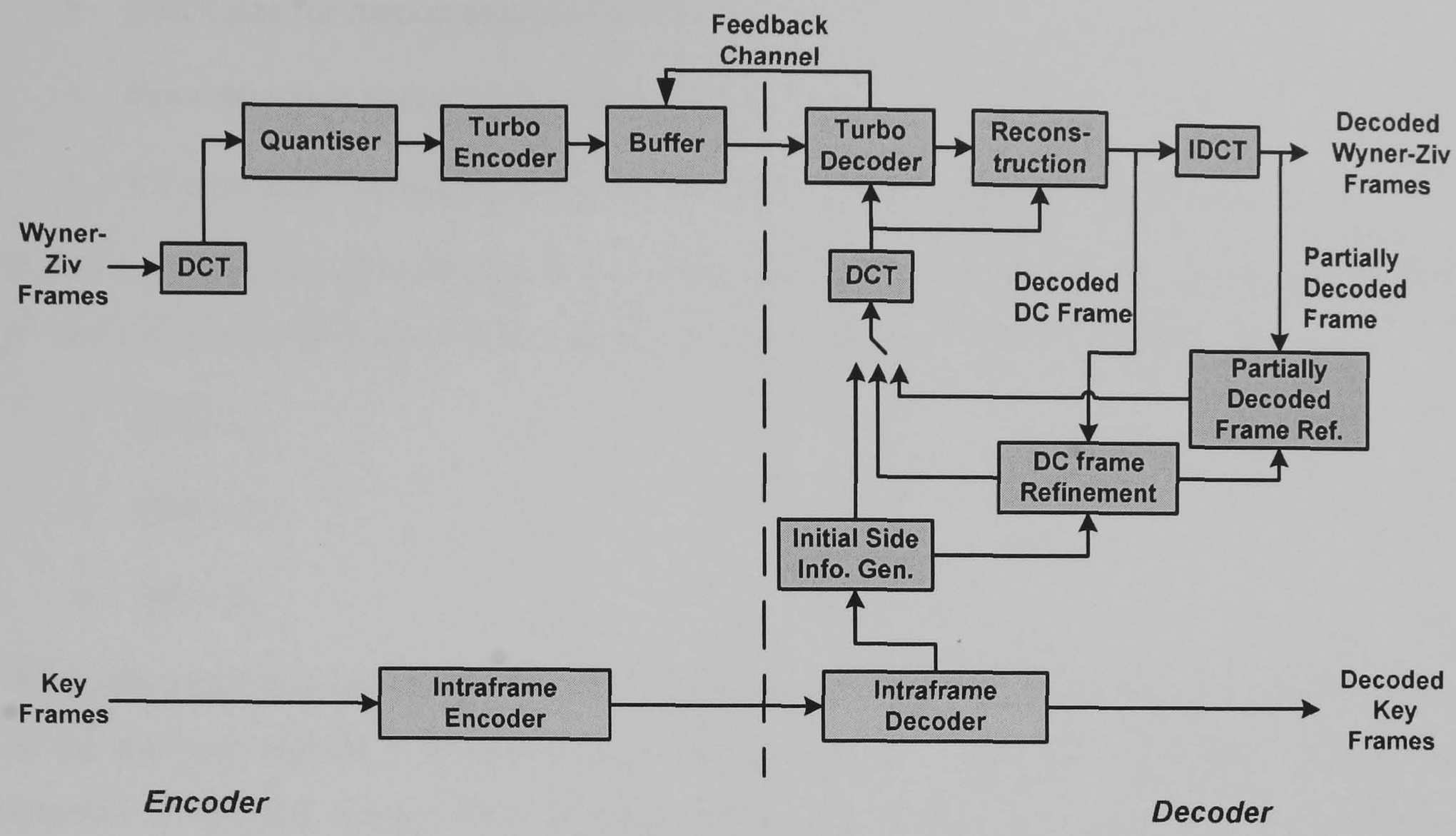


Figure 3-7: Proposed iterative SI refinement codec architecture

The new refined side information is synthesised considering forward, backward, bi-directional predictions and also the motion vectors from the previous refinement of the side information. The mode and the vectors that minimise the SAD are selected for the synthesis of SI.

Proposed refinement technique strongly relies on the decoded information, hence the errors in the turbo decoding process is unrecoverable. The effect is stronger if there are errors in DC coefficients. Therefore, for DC coefficients a reduced Bit Error Rate (BER) threshold can be used for the proposed technique in order to minimise this effect.

3.4.2 Experimental Results

In the experiments *Foreman*, *Soccer*, *Hall-Monitor* and *Coastguard* QCIF sequences are considered as in the previous section. The bit rate and PSNR are calculated for the luminance component of all frames by averaging them over the sequence. Key frames are intra-coded using H.264/AVC and quantisation parameters are selected to match the average PSNR of WZ frames and key frames. The quantisation matrices given in Figure 2-29 (page 42) are considered, to obtain different rate-distortion points. Other test conditions are:

- Frame rate: 15 fps
- GOP length: 2, 4 and 8
- DCT size: 4×4

- Block size for motion estimation: 8×8
- Error detection mode: Ideal error detection
- Bit error rate threshold: 0 for DC coefficients, 10^{-3} for the other coefficients

Figure 3-8 shows the refinement of side information starting from the initial side information. 2nd, 8th and 12th frames of *Soccer* QCIF sequence for the following additional test conditions:

- GOP: 2
- QM = 3
- QP = 41

When the initial side information and DC refined side information (the proposed technique given in the previous section - Section 3.3) compared, there are major improvements, such as the location of moving objects (ball or legs), which can be seen in Figure 3-8. With DC+AC refinement (the proposed technique given in this section), the side information is further refined. In the given frames it can be seen from the shape of the ball, shadows and ground details that more accurate side information is obtained closer to the original frame.

Figure 3-9, Figure 3-10 and Figure 3-11 illustrate the refinement of the side information step by step starting from the initial side information for the *Soccer*, *Foreman* and *Coastguard* sequences in the same order.

- GOP: 2
- QM = 3
- QP = 41 for *Soccer*, 38 for *Foreman*, 37 for *Coastguard* sequences

Results show a significant improvement in the quality of SI. The gain is up to 8 dB for the DC refinement and 14.5 dB for the DC+AC refinement for the *Soccer* sequence. The gains are up to 6.2 dB and 9 dB for the *Foreman* and 3.2 dB and 4.5 dB for the *Coastguard* sequences, which show that the proposed side information technique results better for high motion sequences. It is also noticeable that for some frames (higher motion frames within the given video sequences) where the initial side information quality is poor, the gain is more significant than it is in the other frames.

In the results summarised in Figure 3-12 to Figure 3-17 the overall rate-distortion (RD) performance of the proposed side information refinement technique is shown, compared to the DC refinement, H.264/AVC with IBI coding structure, the VISNET-II codec [6] and the refinement technique proposed by Ye et al. [72]. The performance of H.264/AVC is obtained by restricting motion vectors to zero in order to have a similar encoder complexity. A significant

PSNR gain up to 3.5 dB has been achieved by the proposed technique against VISNET-II codec (Figure 3-17). Also the gap between DVC and H.264/AVC is closed for the *Foreman* sequence and decreased for the *Soccer* sequence. As it can be seen in the results, proposed technique improves the rate distortion performance more significantly for high motion sequences than low motion sequences. Moreover the gap between the DC refinement and proposed technique (DC+AC refinement) is more visible for high motion sequences and as the GOP size is increased proposed technique improves the performance more significantly: for the *Foreman* sequence the improvement is approximately 1.5 dB for GOP: 2; 2.5dB for GOP: 4 and 3.5 dB for GOP: 8. For the *Hall-Monitor* and the *Coastguard* sequences the improvement is more visible when the GOP size increased (Figure 3-14 and Figure 3-16).

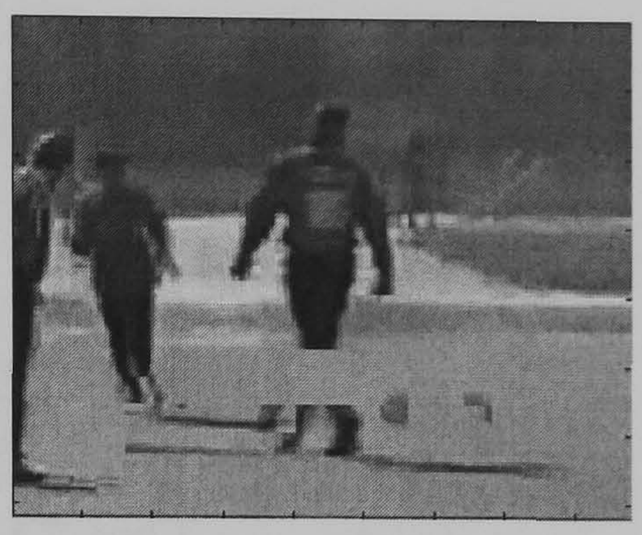
Proposed technique is applied on the decoder only; therefore, encoder complexity is not increased. Average decoding speeds per frame (WZ frame) are given in Table 3.1. In this test, same conditions are provided for VISNET-II codec, DC refinement and the proposed technique (same processor, key frame Quantisation Parameter - QP and WZ Quantisation Matrix - QM). Results show that DC+AC refinement technique decoder complexity is slightly higher except for the *Soccer* sequence and the DC refinement technique complexity is the lowest except for the *Coastguard* sequence. This is because of the simplicity of the operations in the DC domain. However, the DC+AC refinement technique complexity reduces significantly for sequences which have considerable motions as it can be seen from the *Soccer* sequence.

Table 3.1: Decoder speed comparison

Sequence	VISNET-II	DC Ref.	DC+AC Ref.
Coastguard	59.59 s	60.39 s	63.67 s
Foreman	73.31 s	70.79 s	73.78 s
Soccer	92.42 s	84.43 s	86.22 s

(a)

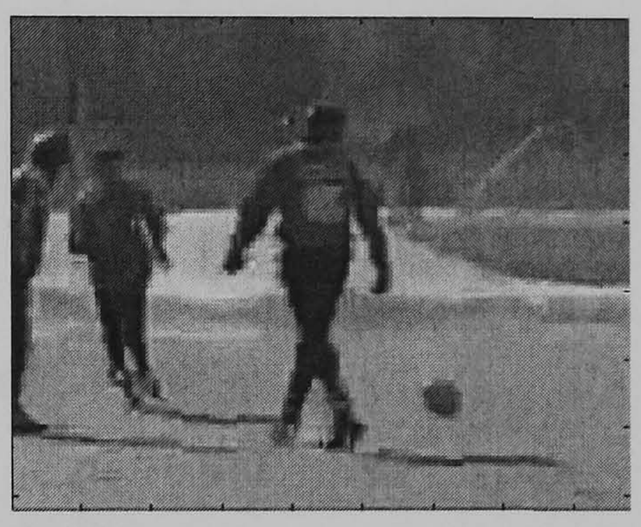
(b)



(c)

(d)

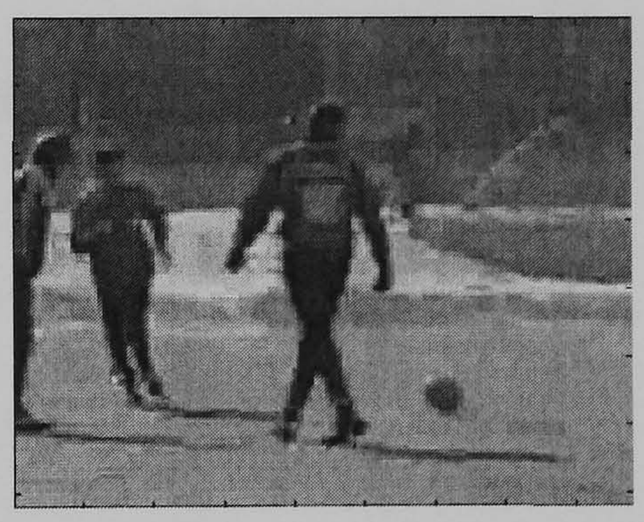
(e)



(f)

(g)

(h)



(i)

(j)

(k)

(l)

Figure 3-8: Side information for 2nd (first column), 8th (second column) and 12th (third column) frames of *Soccer* QCIF sequence; first row-initial SI, second row-DC refined SI, third row-DC+AC refined SI, last row original frames.

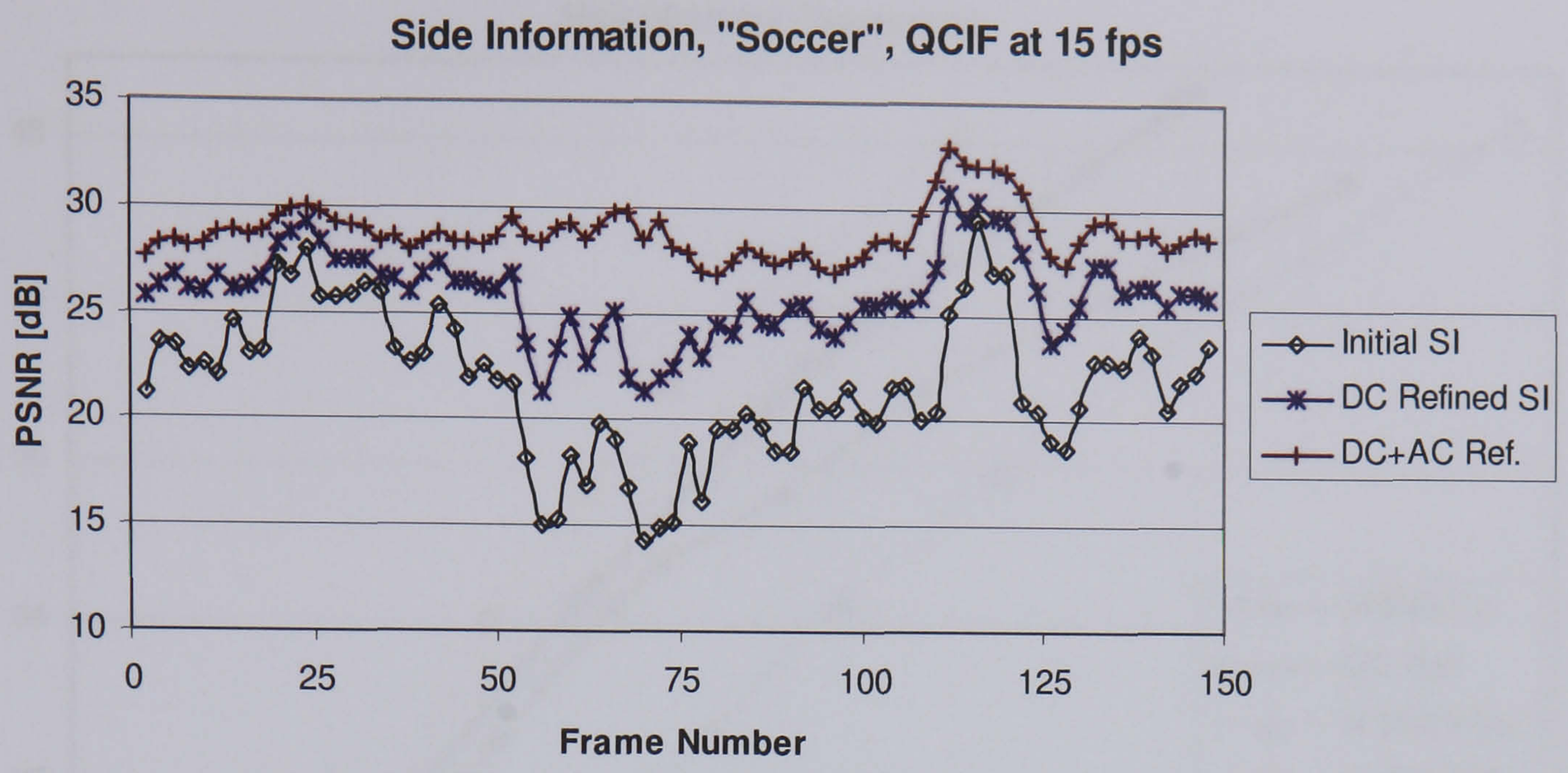


Figure 3-9: Side information comparison for *Soccer* sequence

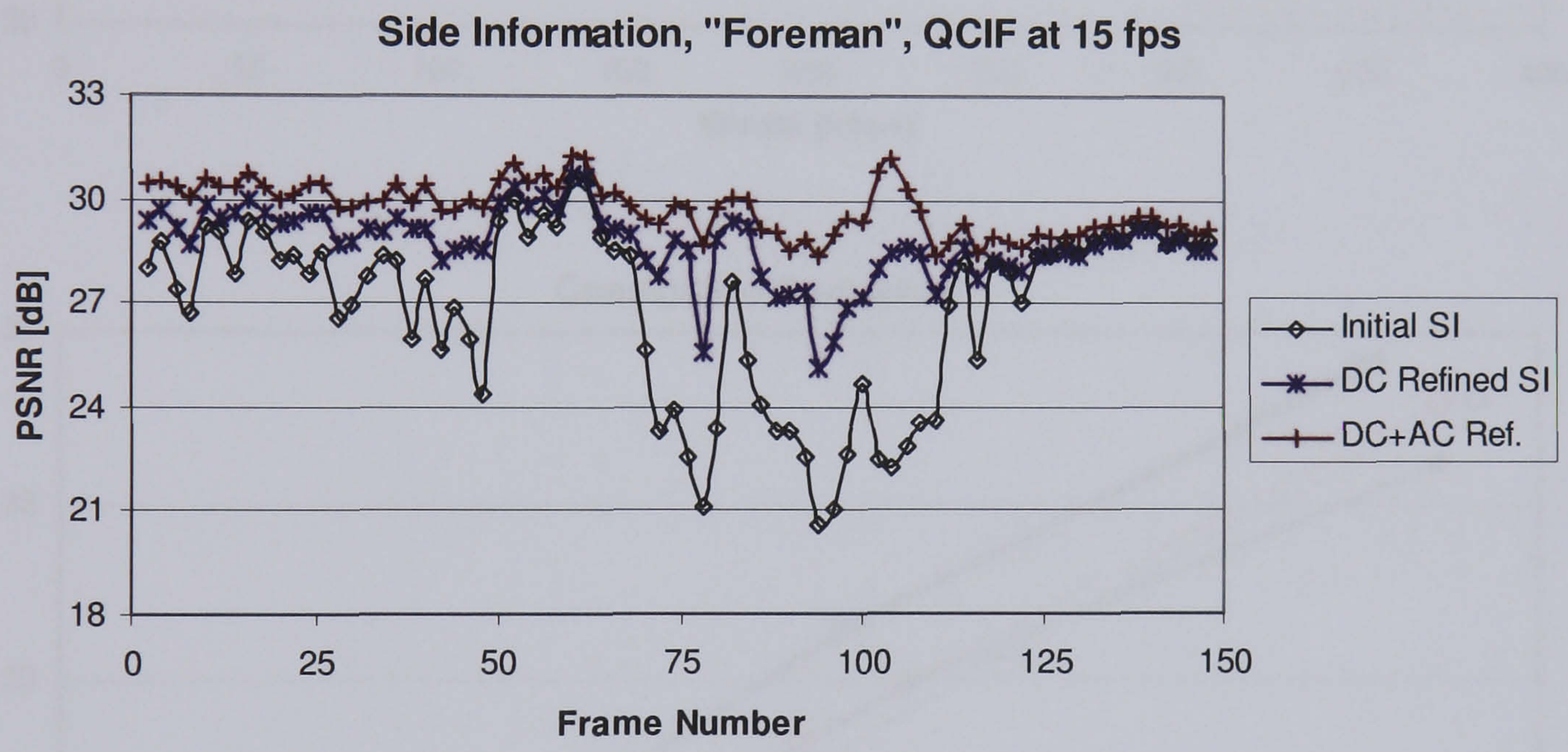


Figure 3-10: Side information comparison for *Foreman* sequence

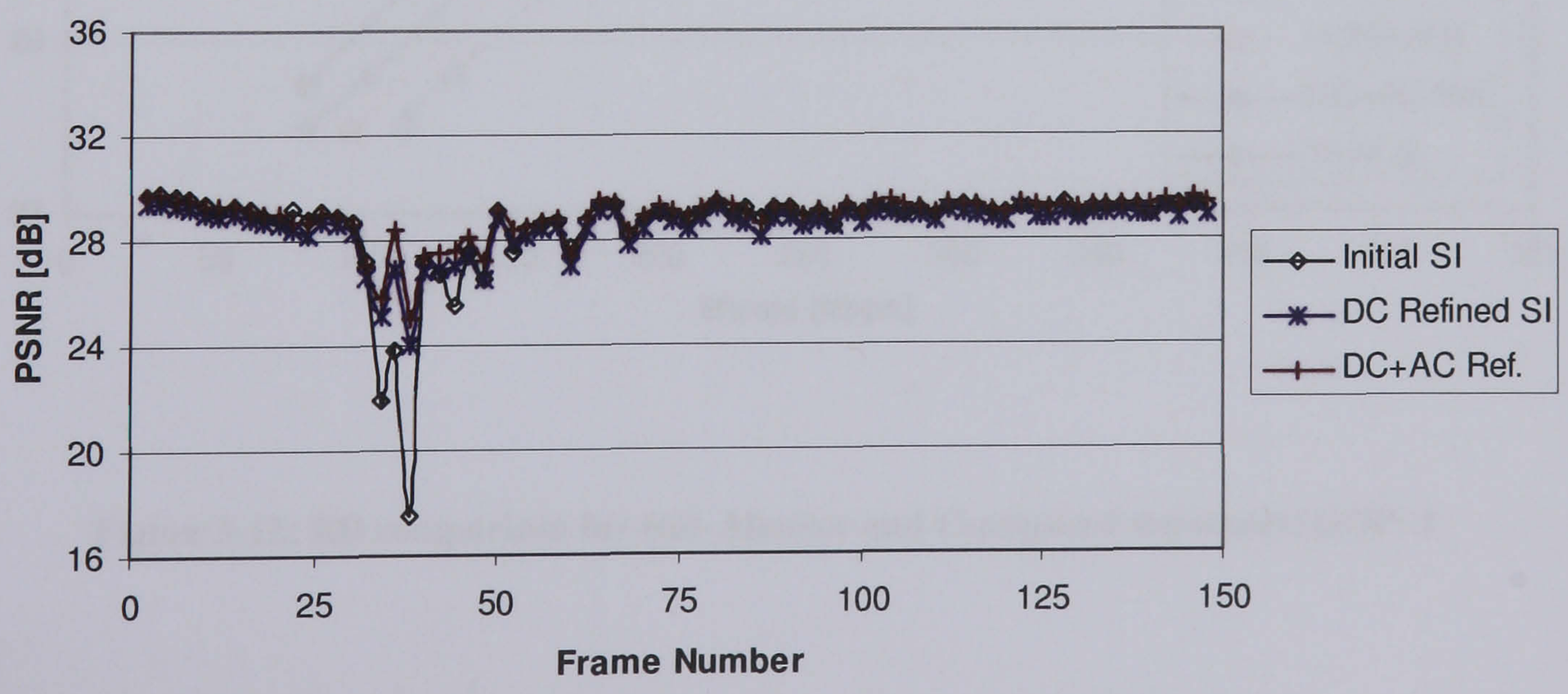
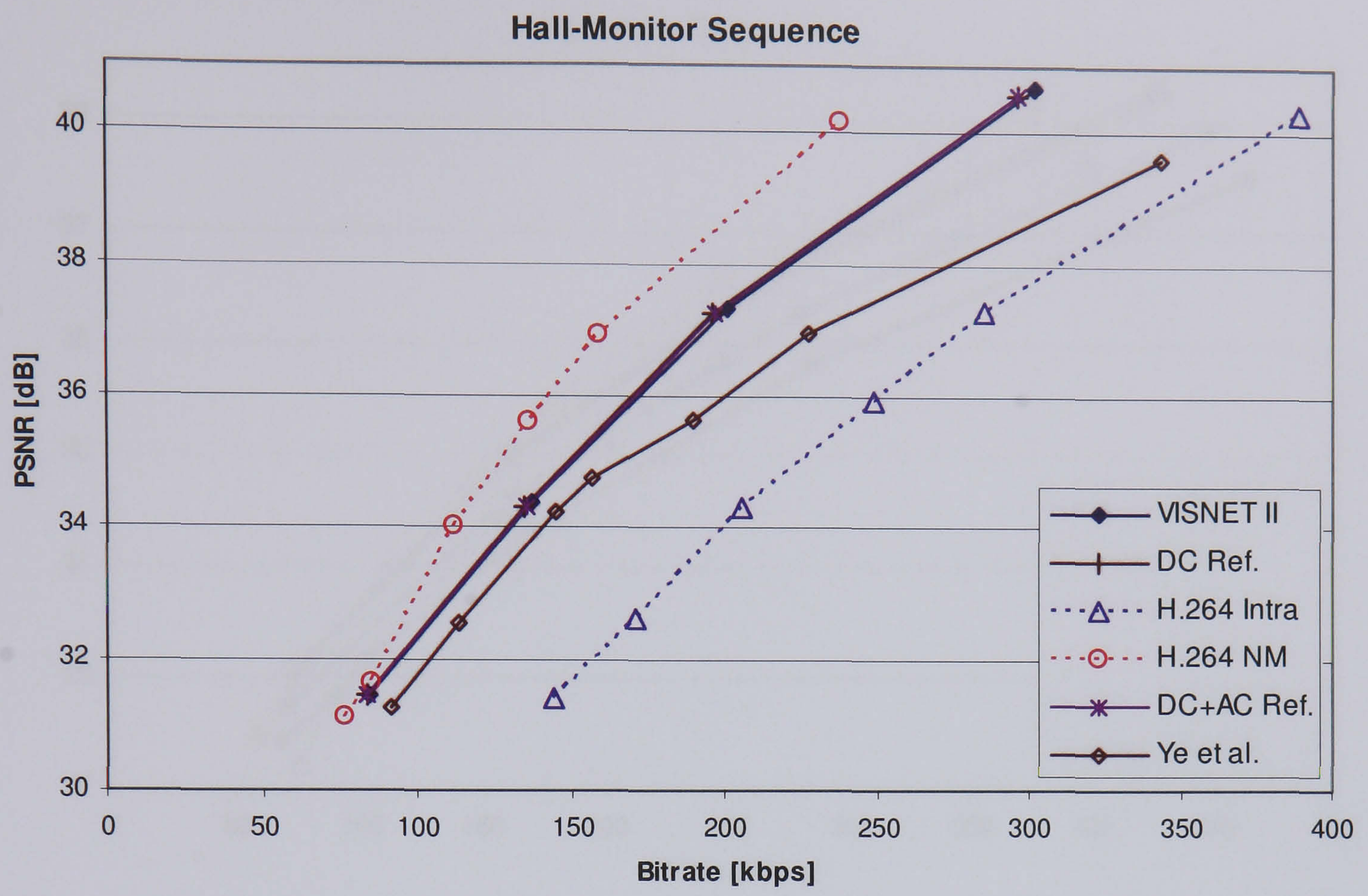


Figure 3-11: Side information comparison for *Coastguard* sequence



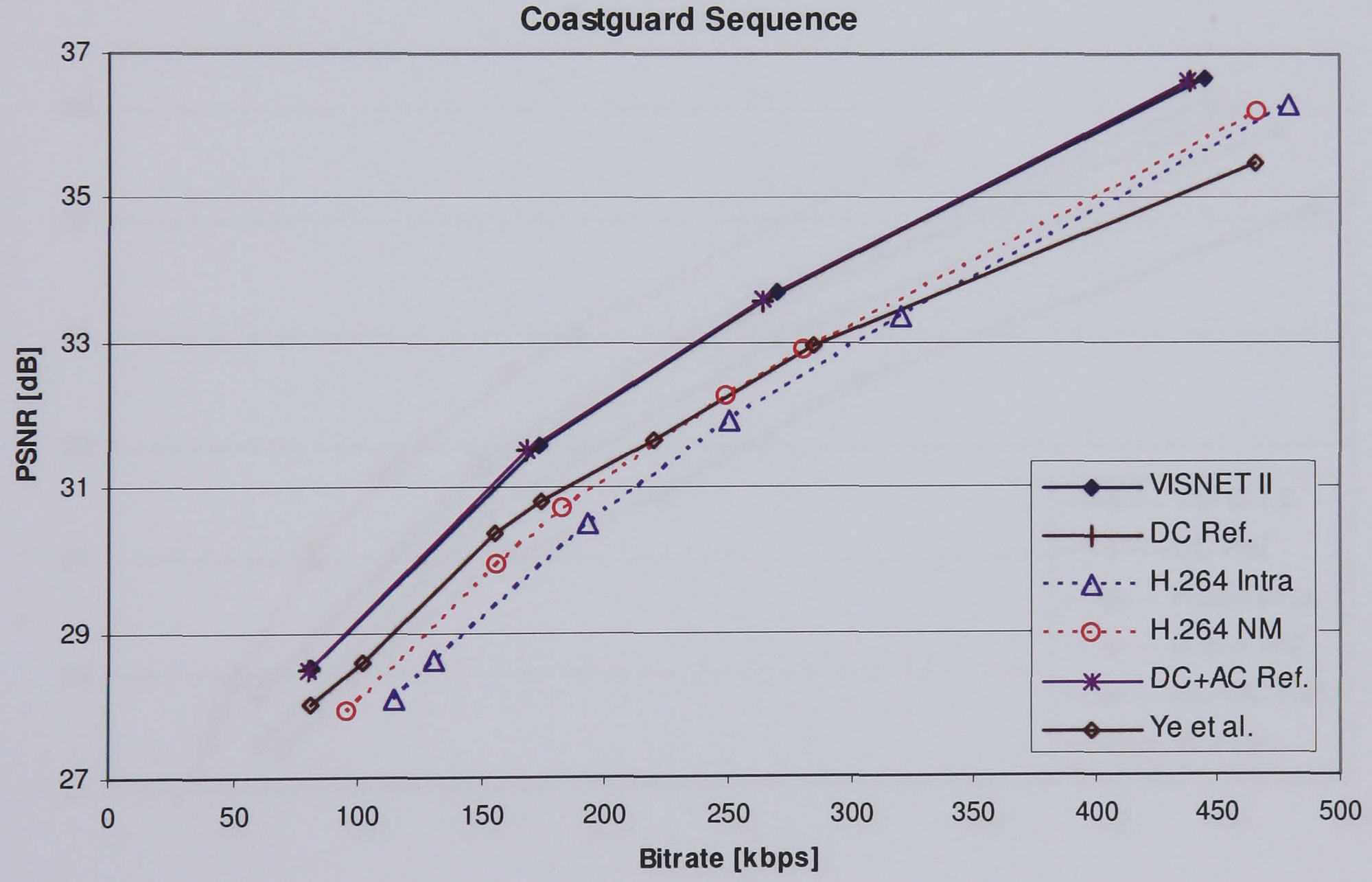


Figure 3-12: RD comparison for *Hal- Monitor* and *Coastguard* sequences; GOP: 2

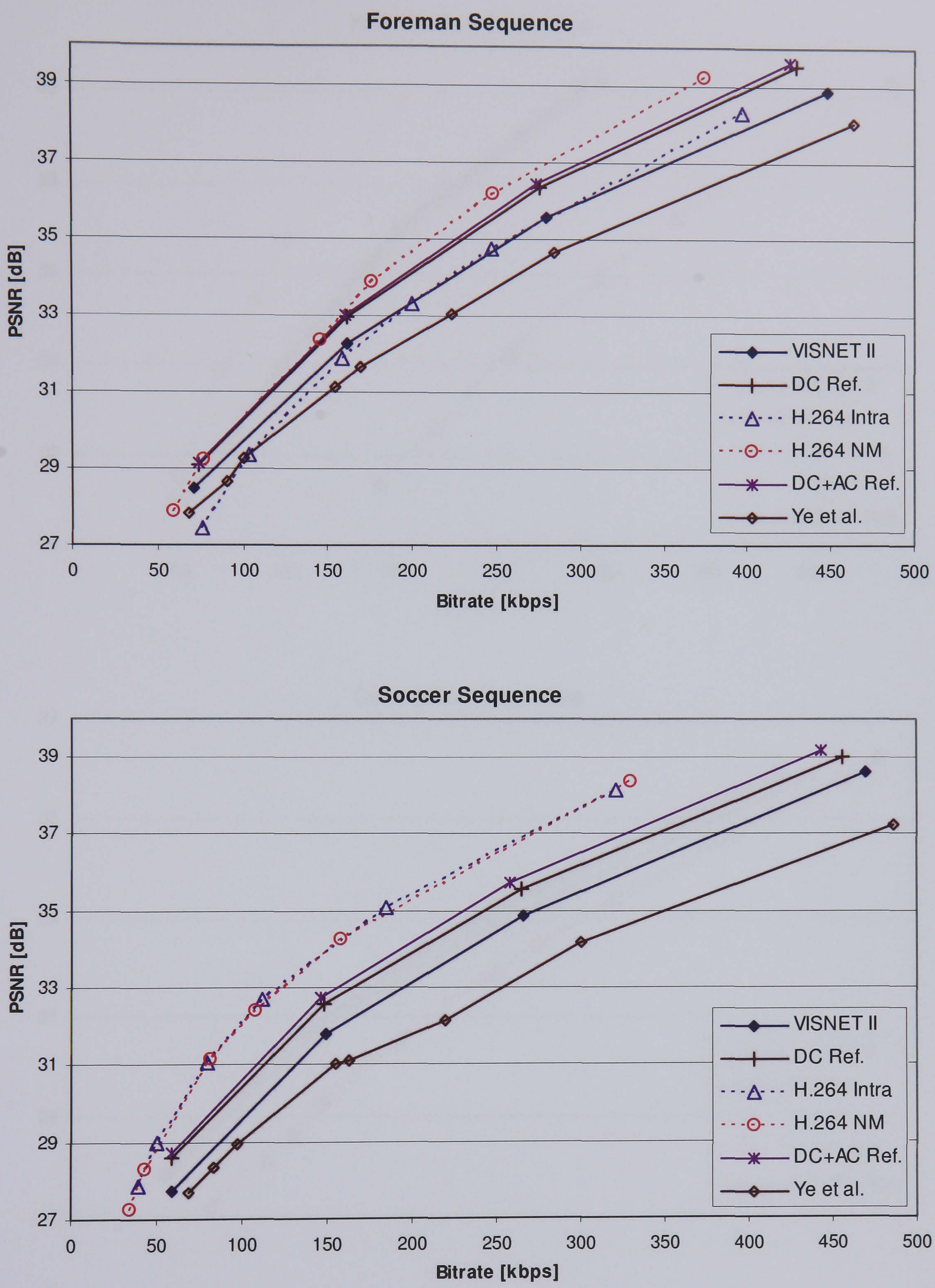
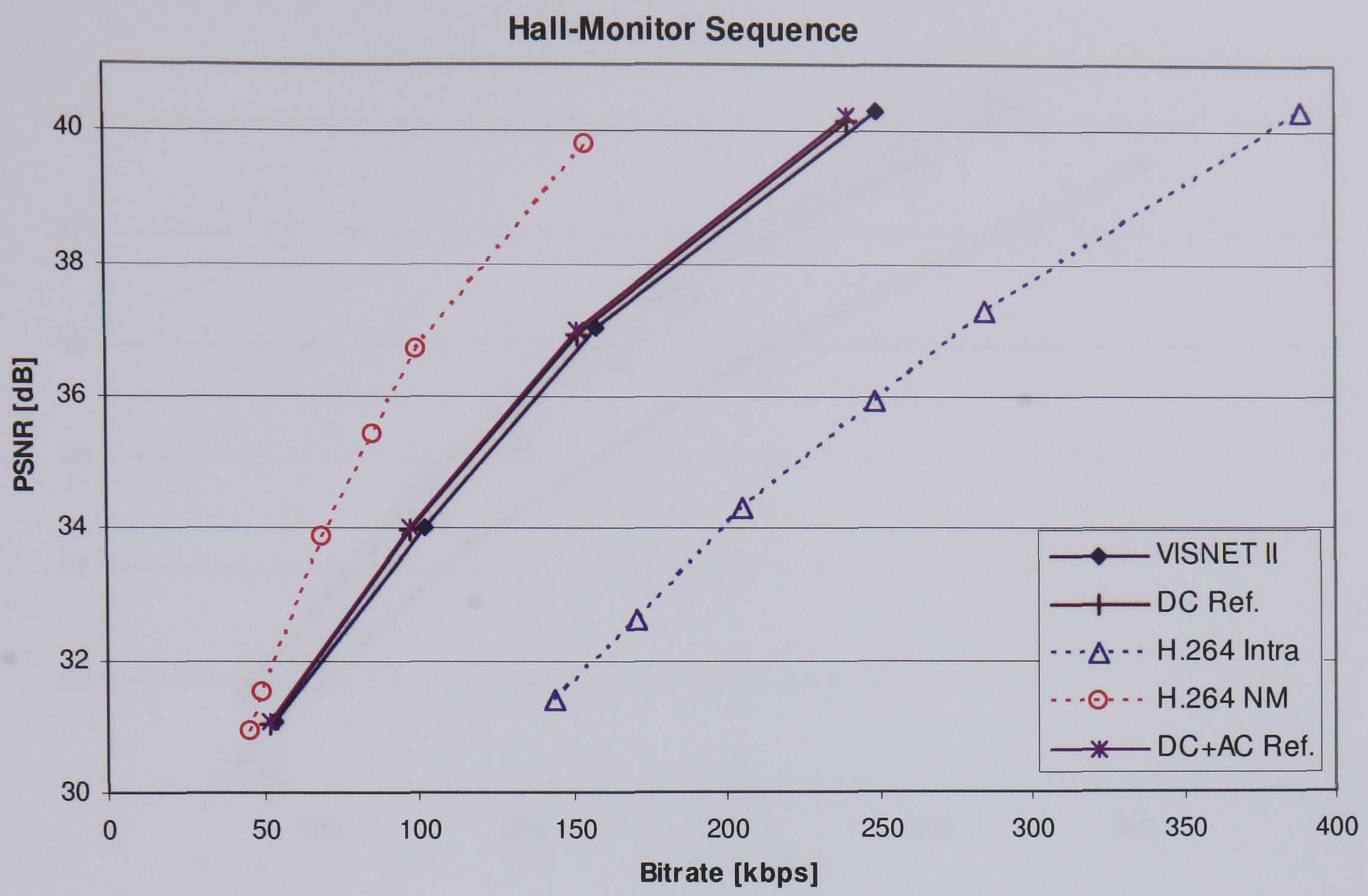


Figure 3-13: RD comparison for *Foreman* and *Soccer* sequences; GOP: 2



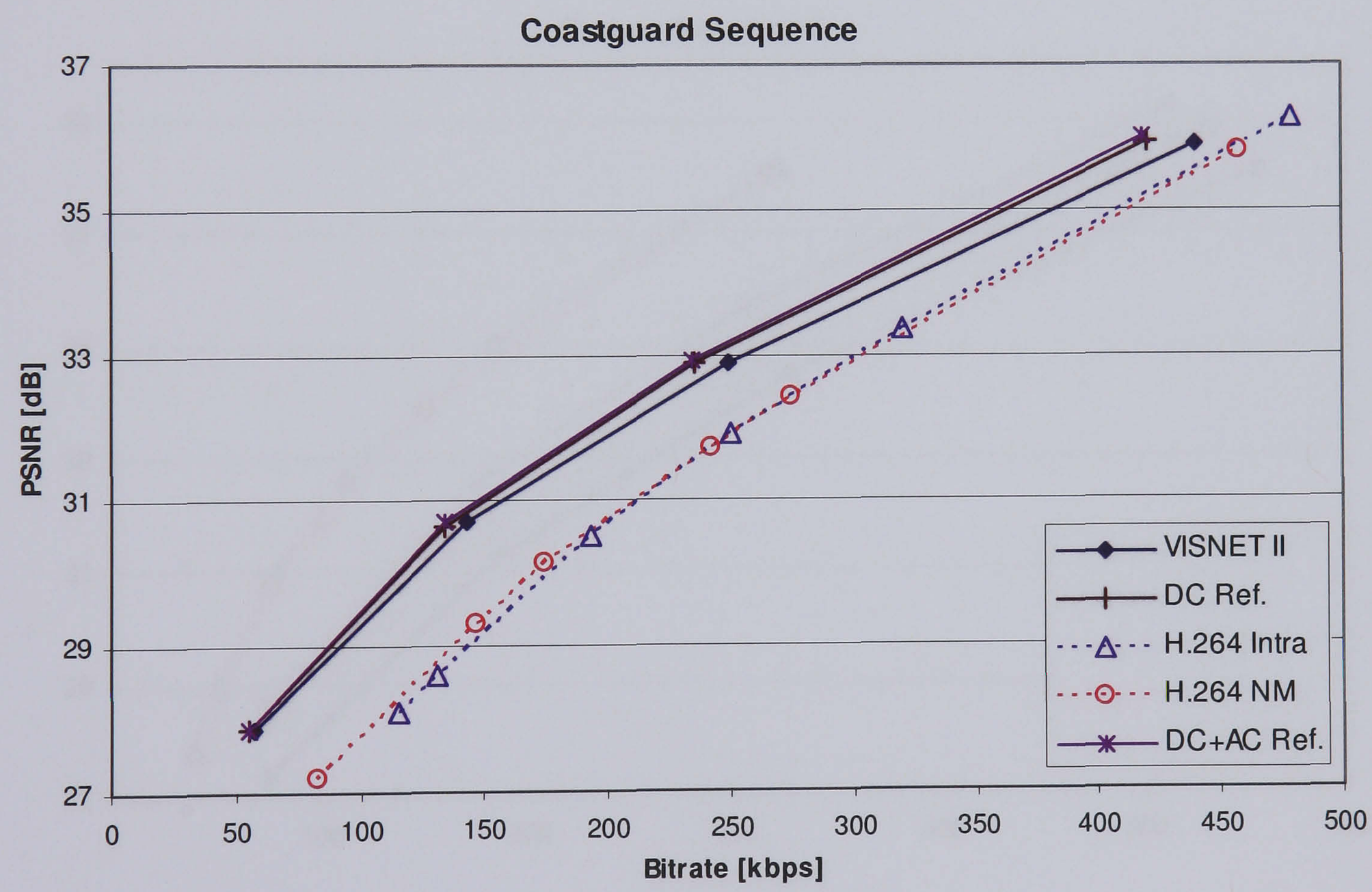
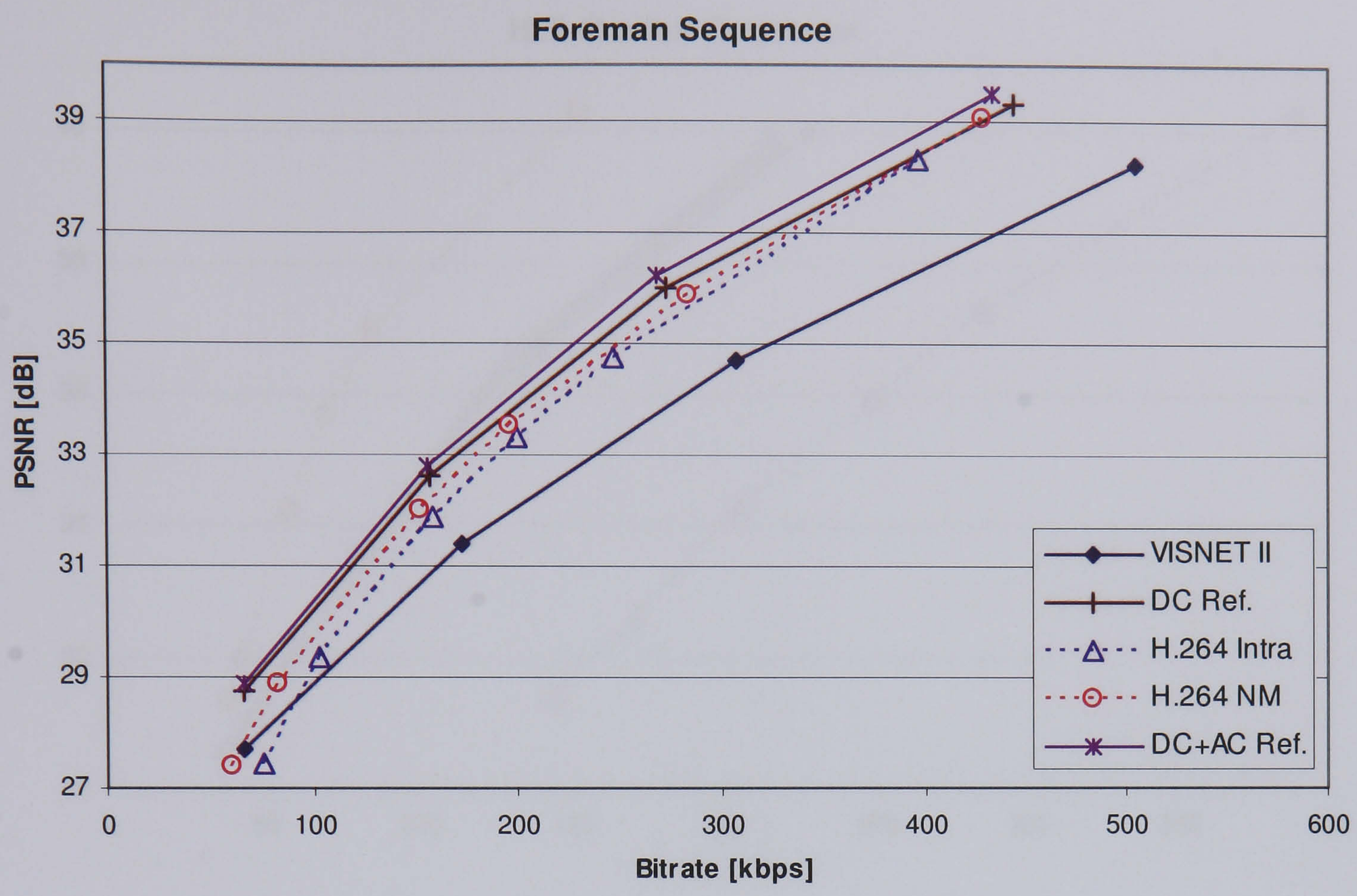


Figure 3-14: RD comparison for *Hall-Monitor* and *Coastguard* sequences; GOP: 4



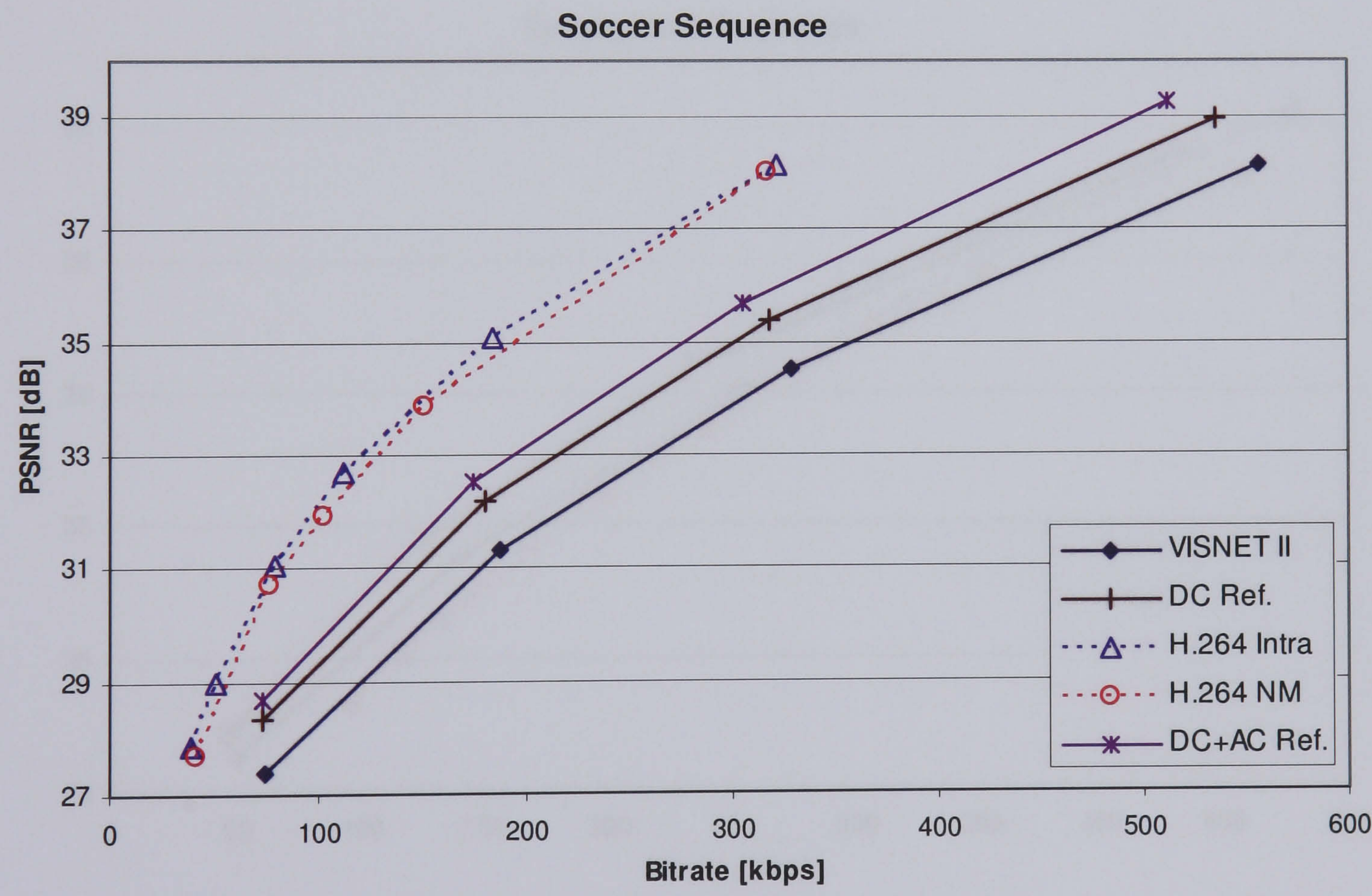
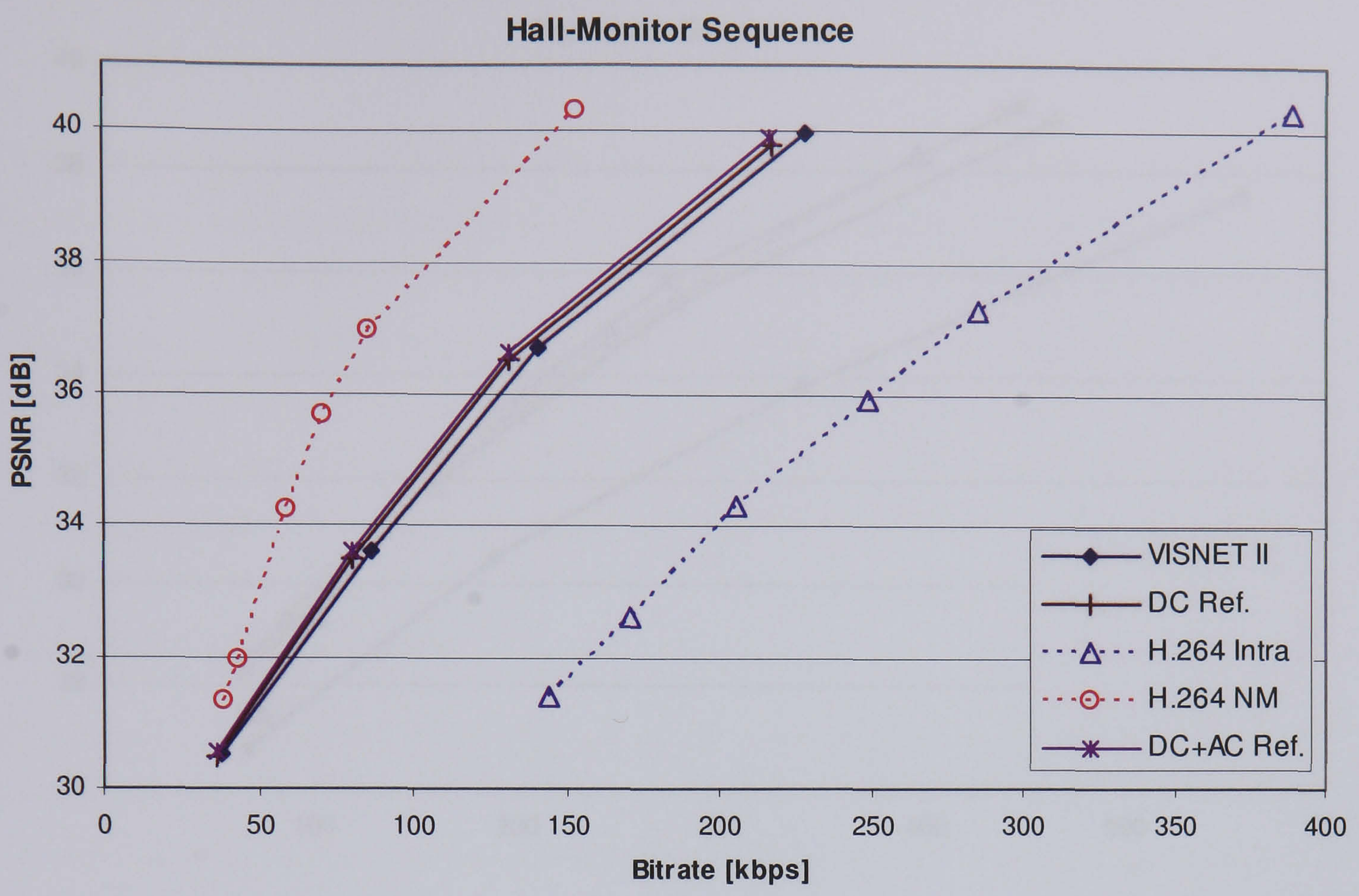


Figure 3-15: RD comparison for *Foreman* and *Soccer* sequences; GOP: 4



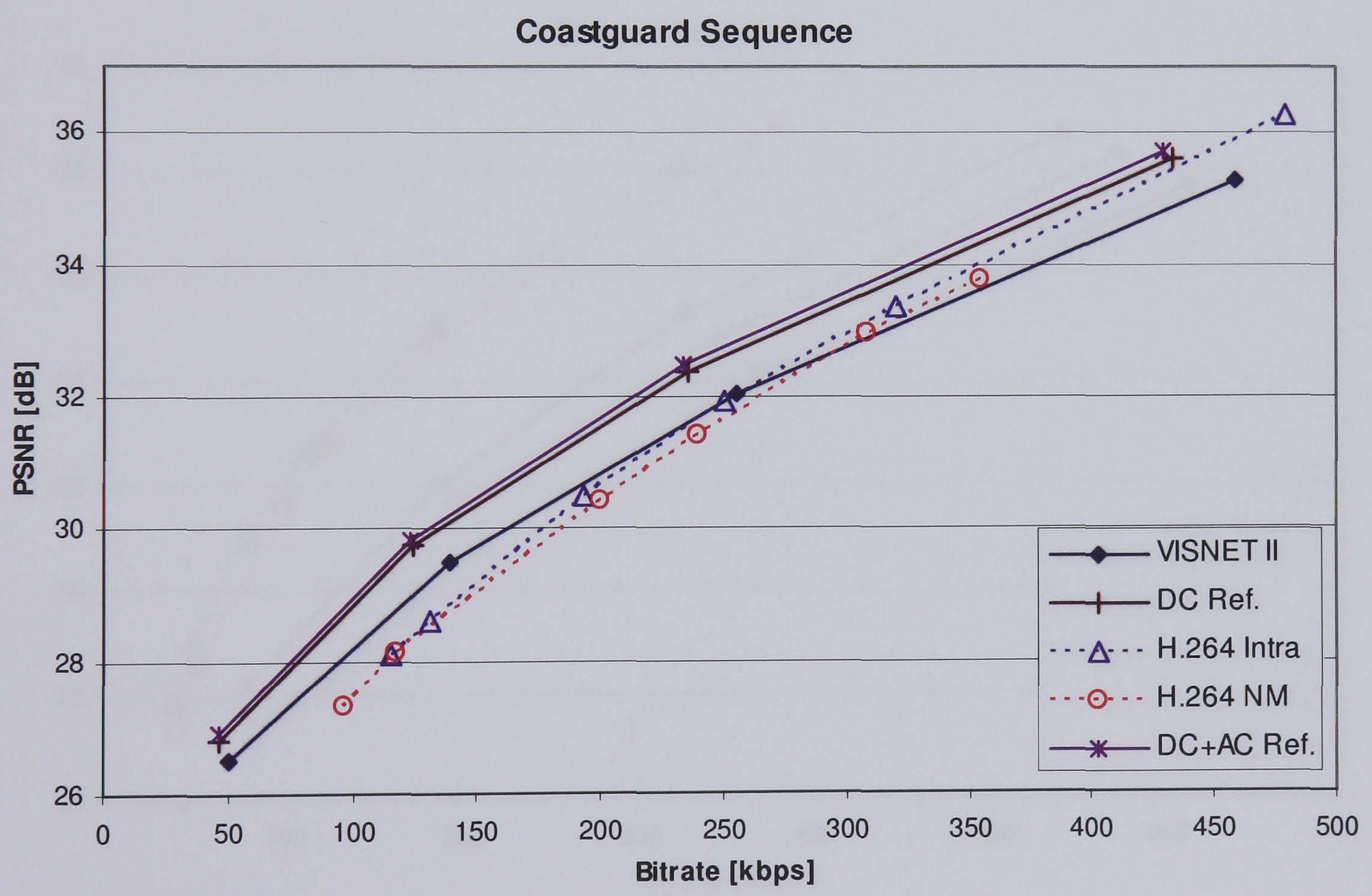
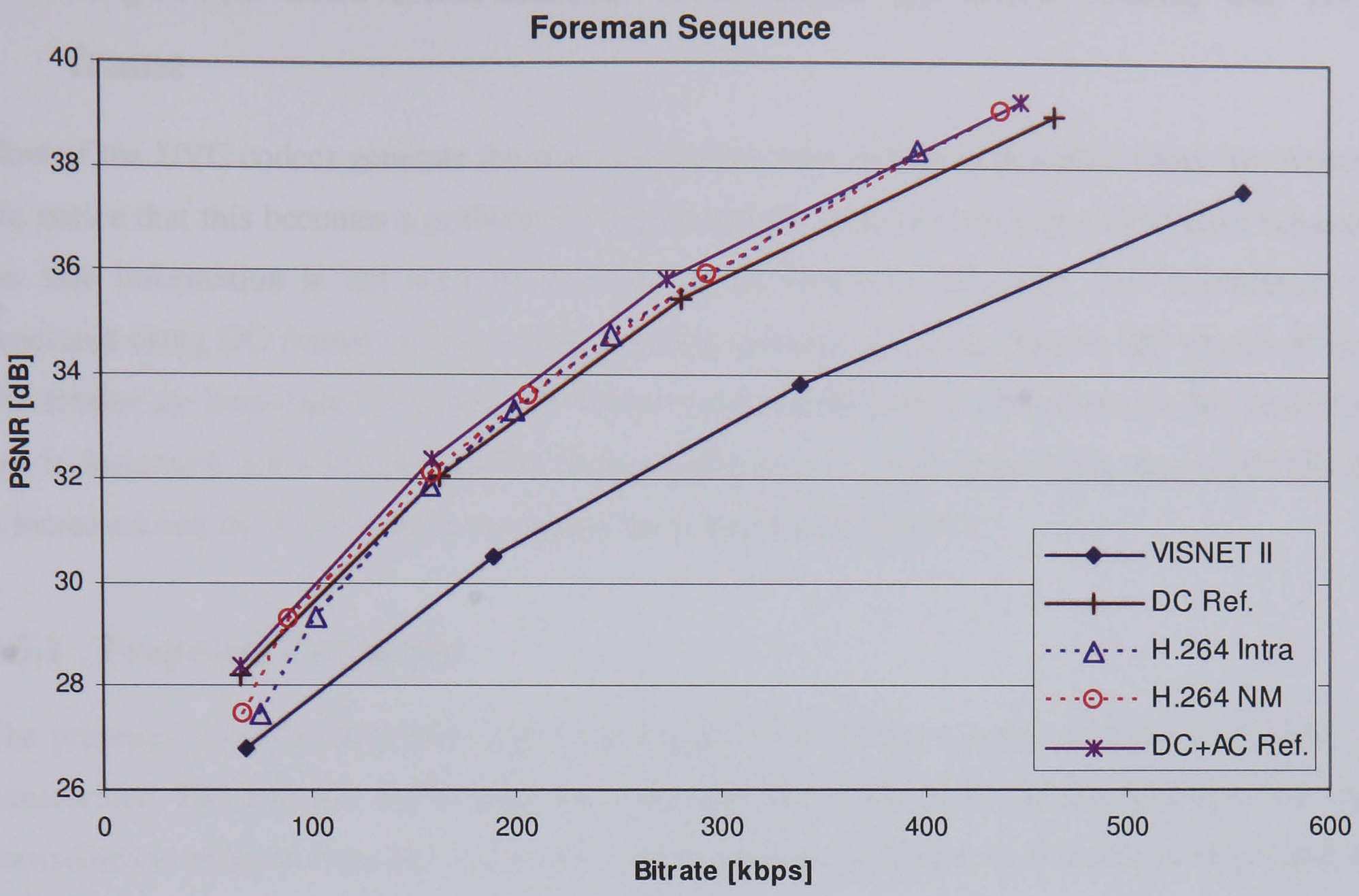


Figure 3-16: RD comparison for *Hall-Monitor* and *Coastguard* sequences; GOP: 8



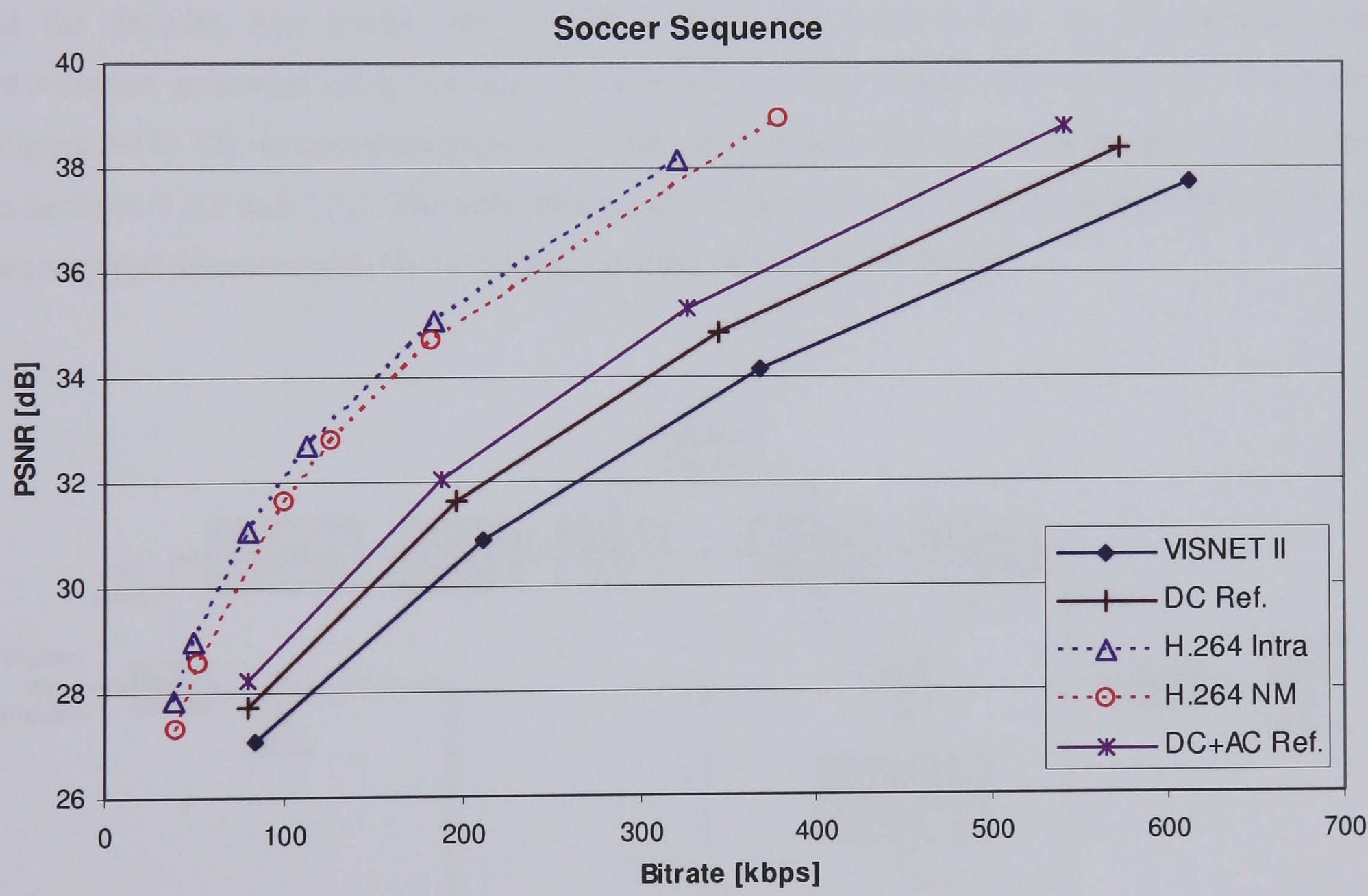


Figure 3-17: RD comparison for *Foreman* and *Soccer* sequences; GOP: 8

3.5 Improved Side Information Generation by Intra-coding the DC frame

Most of the DVC codecs generate the side information with lack of information about that frame. We notice that this becomes a problem for higher motion sequences or larger GOP sizes because the side information is not accurate enough. In the proposed technique, side information is generated using DC frames together with previous and next reference frames. DC frames of the WZ frames are intra-coded to obtain less bit rate for the same quality. Therefore, the DC frame bit rate is decreased, while keeping the DC frame quality as it is. At the same time, the overall PSNR is increased and the AC bit rate is decreased (so is the overall bit rate).

3.5.1 Proposed Technique

The proposed codec architecture is given in Figure 3-18. At the encoder after the WZ frame is transformed, DC and AC coefficients are separated. DC coefficients are grouped together and normalised in order to form DC frame. AC coefficients are quantised, turbo encoded and stored in the buffer.

At the decoder, key frames and DC-WZ frames are intra-decoded. At the decoder, side information generated using decoded previous and next key frames and current DC WZ frame (Figure 3-19). DC motion estimation and mixed mode motion compensation are used as described in Sections 3.3.1 and 3.3.2. The side information is then DCT transformed, and fed into the turbo decoder and reconstruction block for the decoding process of AC bands.

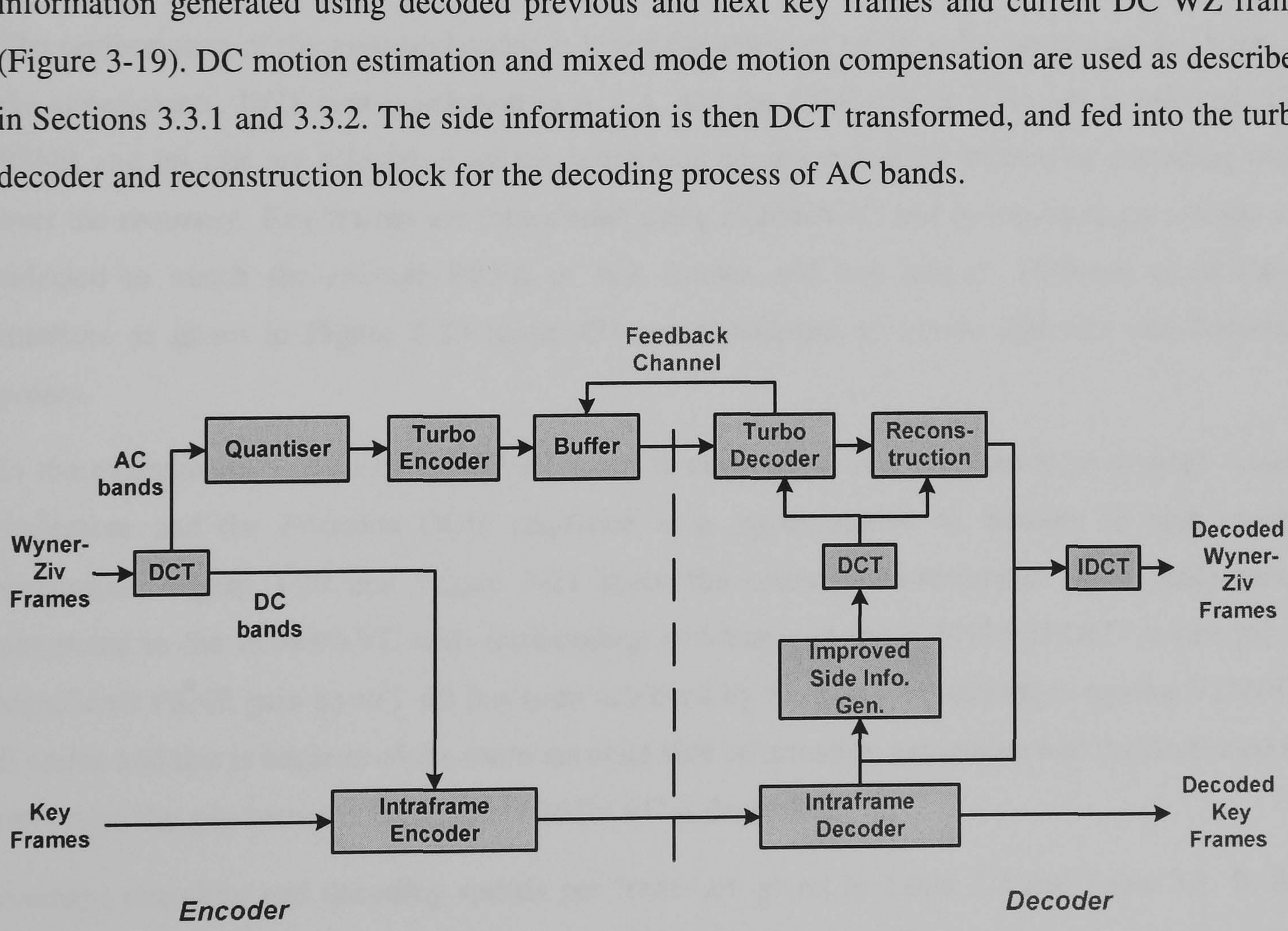


Figure 3-18: Proposed codec architecture

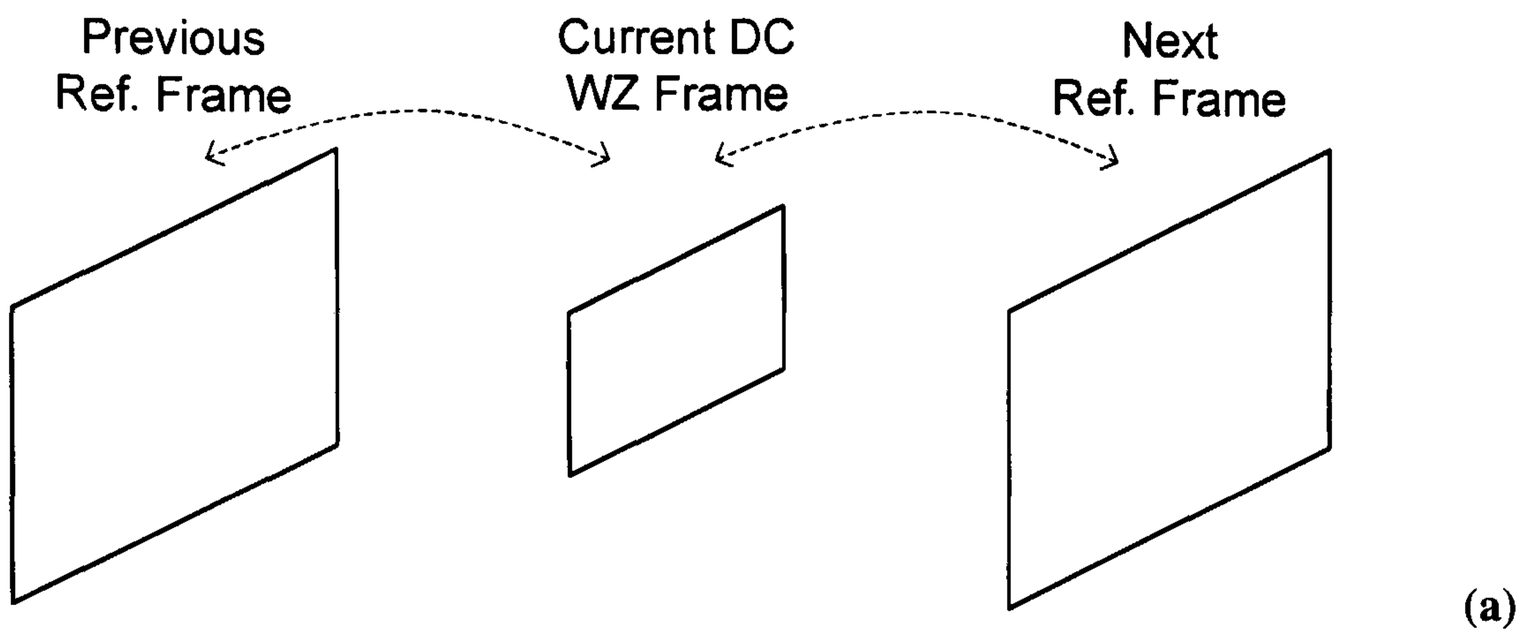


Figure 3-19: Side information generation (a) Motion estimation and (b) Side information synthesis

3.5.2 Experimental Results

The performance of the proposed codec is tested for standard QCIF video sequences at 15 fps. In the experiments, DCT size is selected as 4 × 4, and the GOP size of 8 frames is selected. The PSNR and bit rate are calculated for the luminance component of all frames by averaging them over the sequence. Key frames are intra-coded using H.264/AVC and quantisation parameters are selected to match the average PSNR of WZ frames and key frames. Different quantisation matrices as given in Figure 2-29 (page 42) are considered, to obtain different rate-distortion points.

In the experiments the *Soccer* QCIF sequence is considered as a representative of high motion sequences and the *Foreman* QCIF sequence as a representative of medium to high motion sequences. Figure 3-20 and Figure 3-21 show the overall rate-distortion (RD) performance compared to the H.264/AVC with intra-coding structure and the VISNET-II DCV codec [6]. A significant PSNR gain up to 2 dB has been achieved by the proposed technique against VISNET-II codec and this is because of the more accurate side information generation and the decreased bit rate. Also the gap between DVC and H.264/AVC is decreased.

Average encoding and decoding speeds per frame are given in Table 3.2 and Table 3.3. In this test, same conditions are provided for VISNET-II codec and the proposed technique (same processor, key frames, quantisation parameters and quantisation matrices). Results show that proposed encoder complexity is slightly higher because of the additional intra-coding operations

at the encoder. This increase is acceptable since the WZ encoder has very low complexity operations and any modification will drastically increase encoding times. However, the decoder complexity is significantly reduced since DC band spends the higher number of iterations [5] and the more accurate side information further decreases the iterations for the other bands.

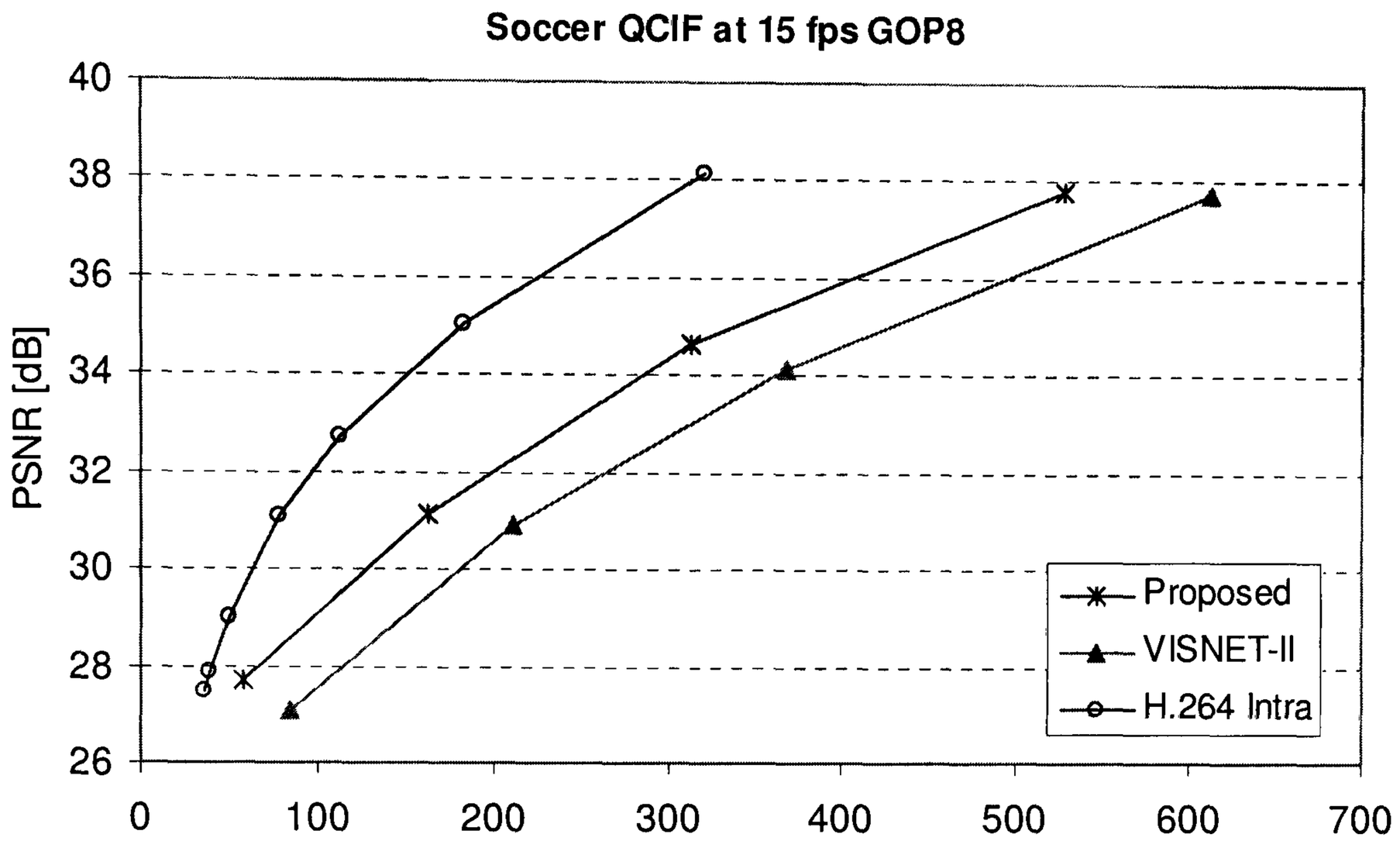


Figure 3-20: Overall RD performances of different codecs for *Soccer* sequence

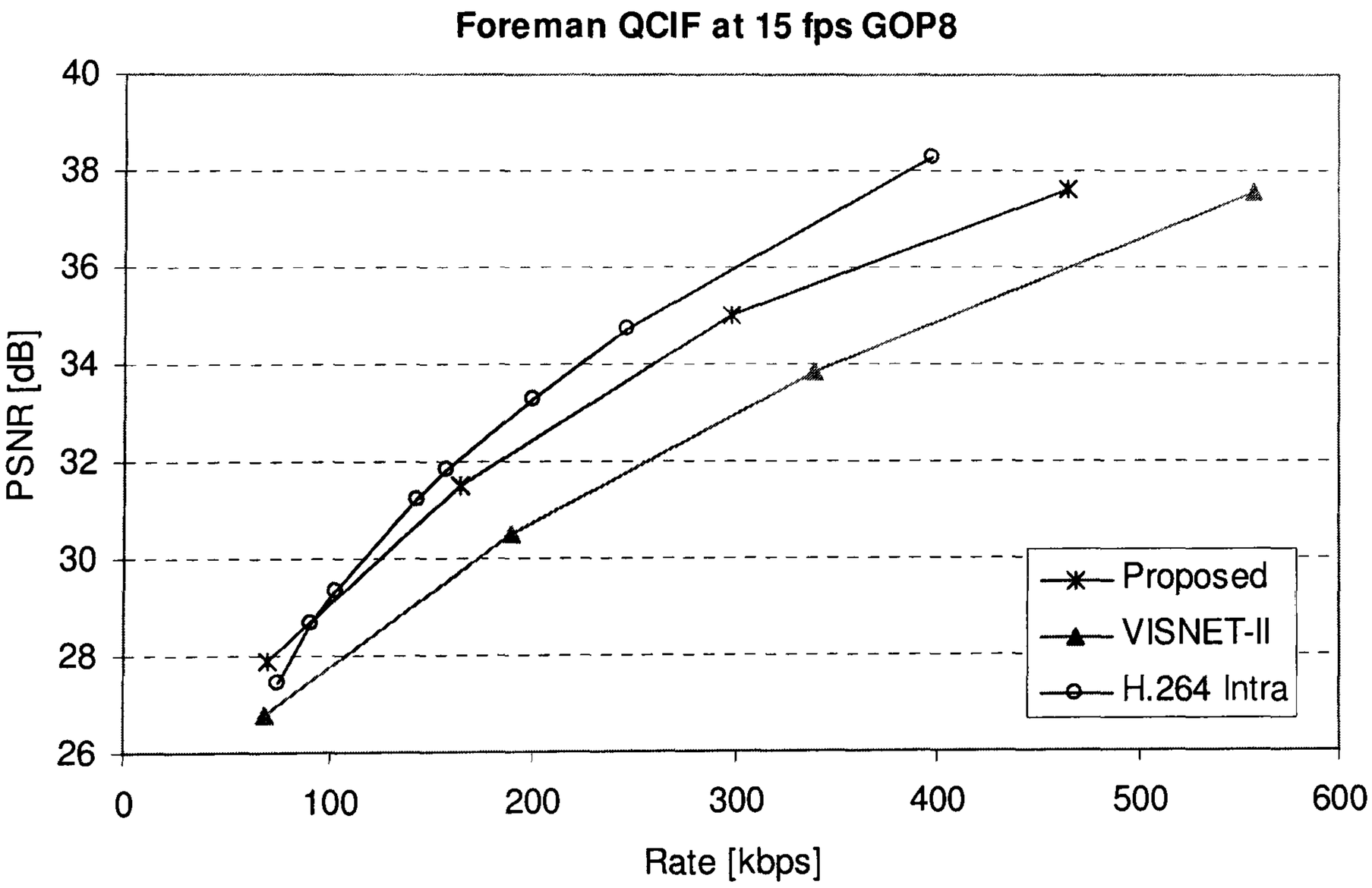


Figure 3-21: Overall RD performances of different codecs for *Foreman* sequence

Table 3.2: Encoder speed comparison

Sequence	VISNET-II	DC Intra-coding
Foreman	42.63 ms	55.30 ms
Soccer	41.33 ms	52.40 ms

Table 3.3: Decoder speed comparison

Sequence	VISNET-II	DC Intra-coding
Foreman	79.50 s	44.96 s
Soccer	91.77 s	59.12 s

3.6 Conclusion

In this chapter, novel side information improvement techniques have been proposed for transform domain DVC. All the proposed DVC codecs are built on the proposed side information in using the decoded DC frame given in Section 3.2 which makes use of the available information during the decoding process. In the codec that is proposed in Section 3.3, only decoded DC frame is used for refining the side information in the DC domain. Further enhancement of the algorithm is introduced in Section 3.4 where the proposed approach uses partially decoded WZ frame for the other refinement iterations. Simulation results clearly show that the proposed codecs can outperform the VISNET-II and the DISCOVER codec by a significant margin. Furthermore, results suggest that the proposed codec decreases the performance difference between the DVC and H.264/AVC. If the two proposed techniques are compared, the latter (Iterative side information refinement technique) further improves the performance. In all results, it can be clearly seen that proposed techniques achieve better results in higher motion sequences and higher GOP sizes, where the initial side information accuracy is lower. Also it is obvious that the proposed techniques do not increase the encoder complexity as they are implemented into the DVC decoder.

In Section 3.5, another novel side information generation technique is proposed for transform domain DVC based on conventionally intra-coding the DC frame. Again this technique is built on the DC motion refinement algorithm. At the encoder the DC frame is intra-coded and at the

decoder, the decoded DC frame is used for generating the side information together with the reference frames. Simulation results clearly show that the proposed codec can outperform the VISNET-II codec by a significant margin. Furthermore, results suggest that the proposed codec closes the performance gap between the DVC and H.264/AVC.

The proposed side information generation and refinement techniques in this chapter have been published / submitted in conferences or journals (please see appendix B).

Chapter 4

4 Residual Coding and Nonlinear Quantisation

Due to its lightweight encoder architecture, the Distributed Video Coding (DVC) concept has been seen as an attractive alternative to its conventional counterparts for a number of applications. Exceptionally low computational complexity has been achieved by moving redundancy exploitation to the decoder. However, going against this norm, some lightweight redundancy exploitation techniques can be utilised at the encoder end at the expense of a slight increase in computational cost at the encoder. In this chapter, frame difference have been utilised at the encoder in order to practise the above idea.

In the first part of this chapter (section 4.1) a novel residual coding technique is proposed for transform domain (DCT based) DVC. This technique minimises the entropy of a given video frame by taking the pixel-wise difference between the current frame and a reference frame before DCT transformation. Subsequently an improved linear quantisation technique is proposed to take advantage of small transform coefficients.

In section 4.2, a novel non-linear quantisation algorithm is proposed for DVC in order to improve the rate-distortion (RD) performance. The proposed solution is expected to exploit the dominant contribution to the picture quality from the relatively small DCT coefficients when the residual input video signal for the Wyner-Ziv frames is considered in the transform domain. The performance of the proposed solution incorporating the non-linear quantiser is compared with the performance of an existing transform domain DVC solution that uses a linear quantiser.

The rest of this chapter is organised as follows: Proposed residual coding for DVC is presented in section 4.1, proposed non-linear quantisation technique is explained in 4.2 and finally section 4.3 concludes this chapter.

4.1 Residual Coding

In this section a residual DVC codec, which encodes the residual frame with respect to a reference frame, is proposed. In order to keep the complexity of the encoder low, simple frame difference, without motion compensation, is considered for transform domain DVC. Some of the temporal redundancies between the current frame and the reference frame are therefore exploited at the encoder.

Aaron et al. proposed a residual coding scheme using LDPC (Low Density Parity Check) coding to exploit the temporal correlations in the pixel domain [73]. They have also proposed a hash-based side information generation technique to aid the decoder in estimating the motion accurately. In contrast, in the work presented in this section, residual coding is implemented on transform domain DVC codec using turbo codes. Furthermore, the rate distortion performance is optimised by the proposed quantisation technique. The technique and the results presented in this section are published in [74].

4.1.1 Residual DVC Codec Architecture

The architecture of the proposed transform domain Wyner-Ziv residual codec is depicted in Figure 4-1. At the encoder, the difference between the current Wyner-Ziv frame (X_{2i}) and a reference frame is computed. In the proposed technique, the previous key frame (X_{2i-1}) is selected as the reference frame in order to minimise the encoder computation complexity. The previous key frame can easily be obtained from the key-frame encoder. The resulting residual frame is divided into 4×4 blocks which serve as the basic coding units. Subsequently each block is DCT transformed. The DCT coefficients are grouped together according to the coefficient band and subsequently they are quantised.

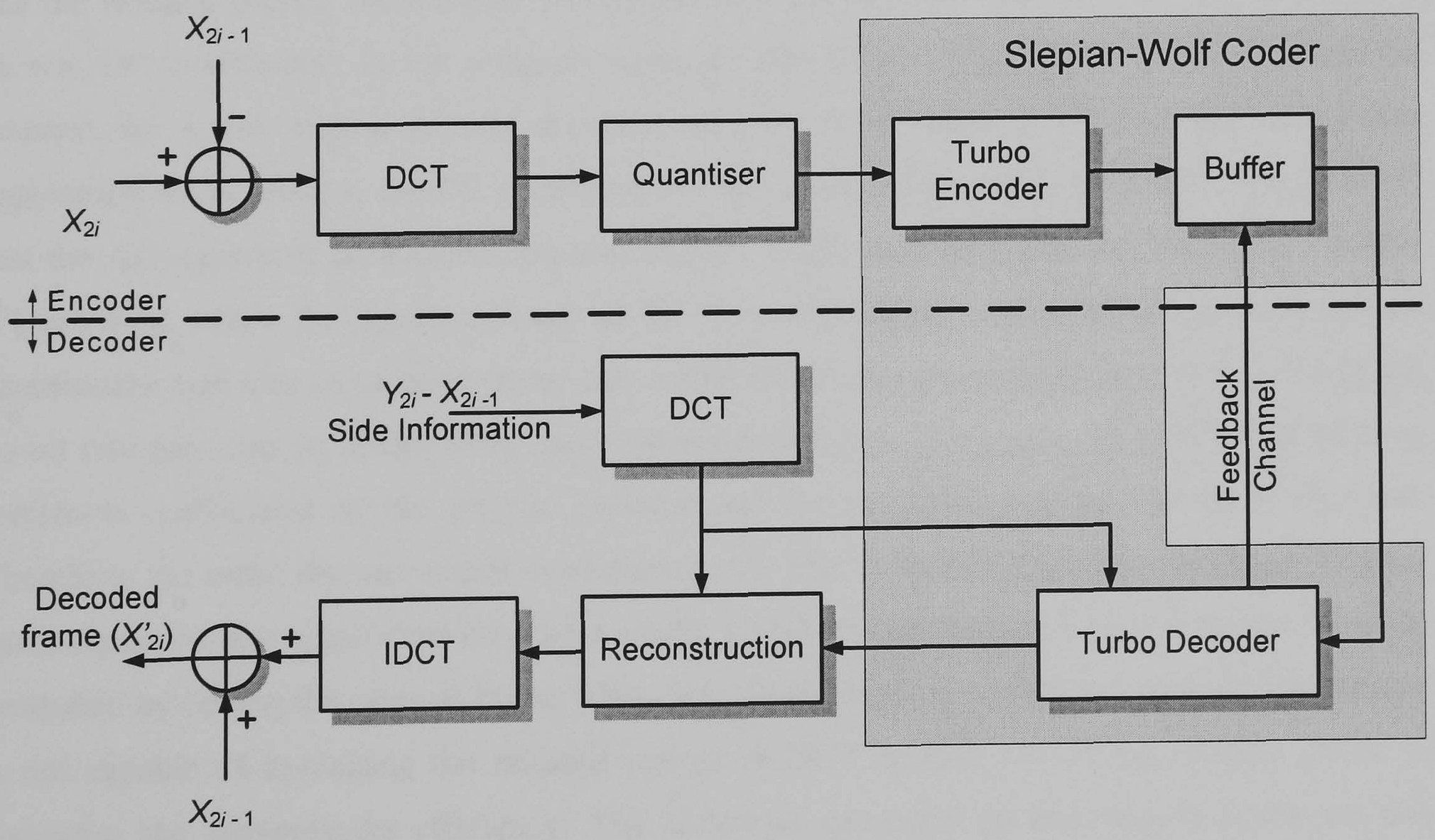


Figure 4-1: Proposed transform domain Wyner-Ziv codec

At the decoder, the prediction of the current frame (Y_{2i}) is generated using the previous and next key frames by the motion interpolation method. The side information is obtained by taking the difference between the interpolated frame (Y_{2i}) and the previous key frame (X_{2i-1}), it is then transformed and fed into the turbo decoder and reconstruction block. After IDCT, the reference frame (previous frame: X_{2i-1}) is added and the decoded frame X'_{2i} is obtained.

In DCT based DVC applications, adaptive quantiser step sizes are used for the AC coefficients [18]. For adaptation purposes, the dynamic range of each coefficient band, which is the maximum absolute value within the frame, is used. The quantisation step size W_k for the DCT band b_k is determined according to Equation (4.1) below:

$$W_k = \frac{2|V_k|_{\max}}{2^{M_k} - 1} \tag{4.1}$$

Where $|V_k|_{max}$ is the highest absolute value within b_k and M_k is the quantisation level value for the DCT band b_k. The decoder receives the dynamic range of each DCT coefficient [18] in order to synchronise encoder and decoder.

4.1.2 Optimising Rate Distortion Performance of Residual DVC

Figure 4-2 illustrates the distribution of DC transform coefficient values (b_0) for the original frame and the residual frame. The average distribution over the first 100 frames of the test sequence is shown. DC coefficients in the original frame are distributed over a large range of values. In contrast, for a low motion activity sequence such as those found in security and surveillance application areas, most of the DC coefficients in the residual frame are very small. It is observed that the AC transform coefficients are also smaller in the residual frame for low motion frames (Figure 4-3). Since the dynamic range of DCT coefficients is small, the use of conventional quantisation step size calculation techniques results in smaller quantisation step sizes. As a result, no bit rate gain can be achieved by using the residual frame. Moreover, the correlation between transform coefficients of the original residual and the predicted residual frames is also less. Therefore, the turbo decoder needs even more parity bits to decode these transform coefficients, particularly the low significant bits. As a result, actual bit rate is even higher than that which is generated by coding the original frame. Thus, it is clear that the conventional quantisation scheme is not capable of exploiting the reduced energy in DCT coefficients of the residual frame to maximise the compression efficiency. The technique presented in this section addresses this problem.

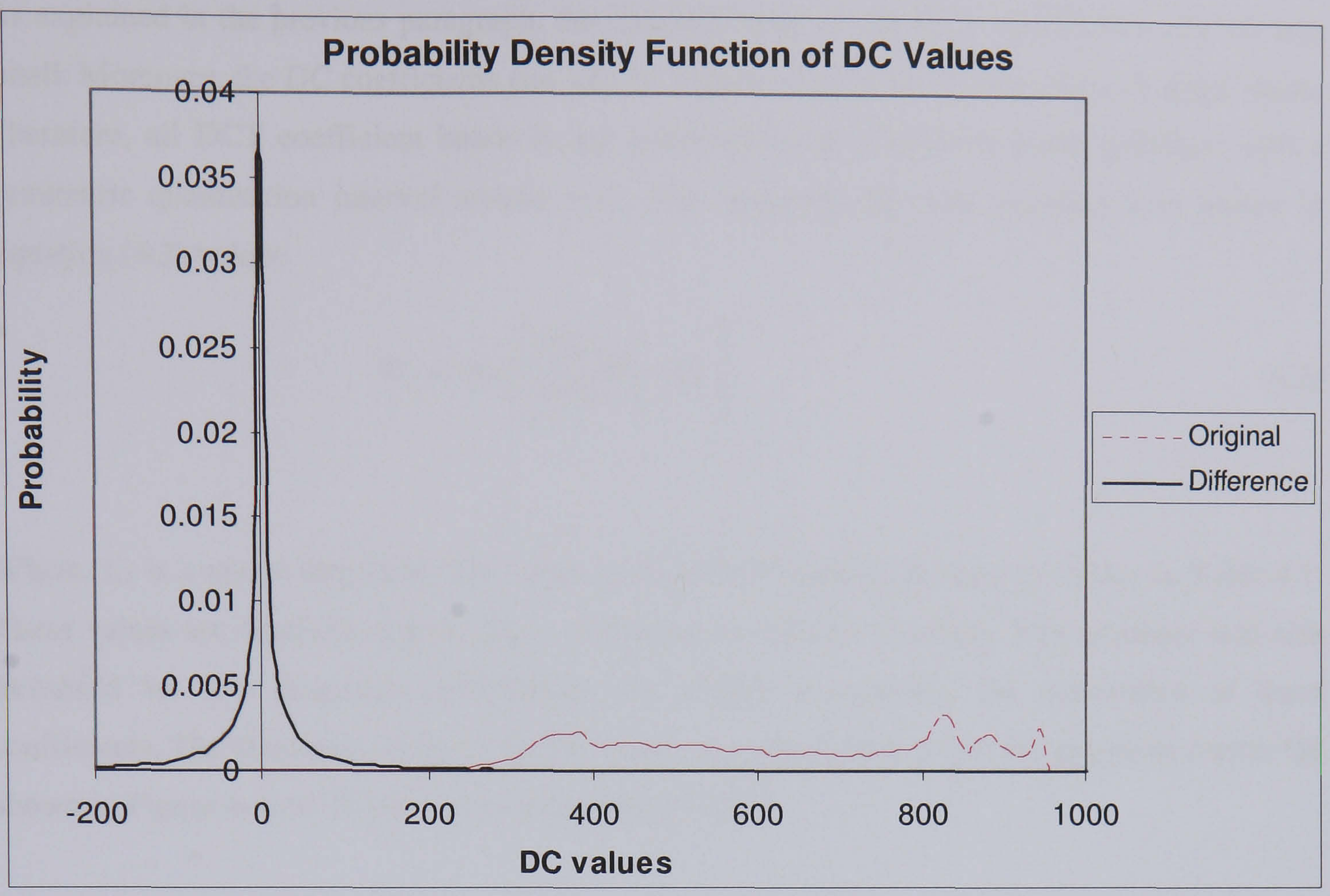


Figure 4-2: Probability density function of DC values for *Foreman* QCIF sequence

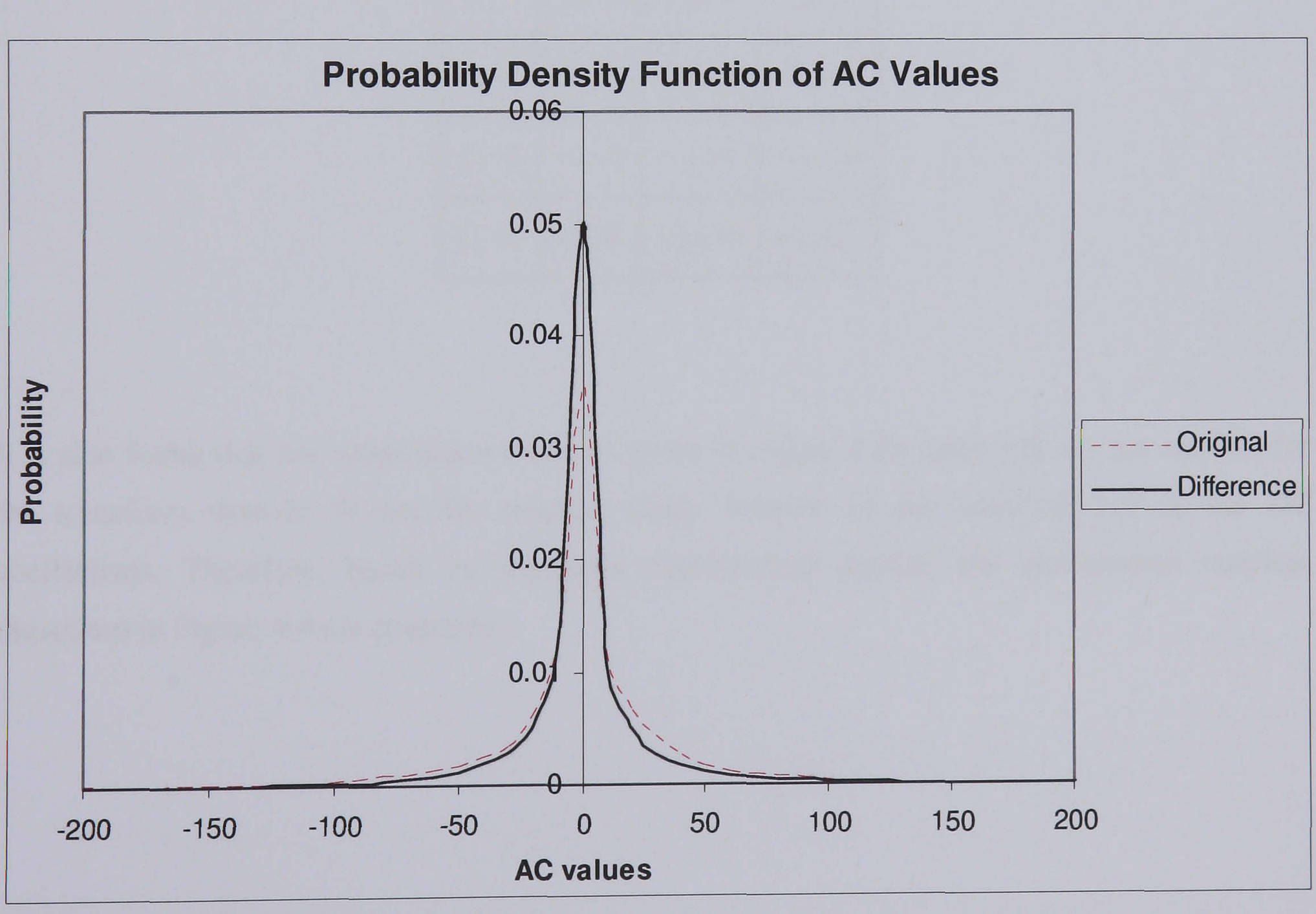


Figure 4-3: Probability density function of AC values for *Foreman* QCIF sequence

As explained in the previous paragraph, the dynamic range of the DCT coefficients may be very small. Moreover, the DC coefficients can also be negative unlike in the case of the original frame. Therefore, all DCT coefficient bands b_k are quantised using a uniform scalar quantiser with a symmetric quantisation interval around zero. The proposed step size calculation is shown in Equation (4.2) below:

$$W_k = \max\left(\frac{2|V_k|_{\max}}{2^{M_k}-1}, C_k\right) \tag{4.2}$$

Where, C_k is a preset threshold. The value of C_k is DCT band dependant as shown in Table 4.1. These values are obtained empirically to maximise the objective quality. The quantiser step size threshold for low frequency coefficients are smaller recognizing the importance of these coefficients. The threshold value of the DC band, C_0, ranges from 6 (for the quantiser matrix Q8 shown in Figure 4-4) to 16 (for the quantiser matrix Q1).

Table 4.1: Minimum step size values for DCT coefficients

C_0	C_0+2	C_0+4	C_0+6
C_0+2	C_0+4	C_0+6	C_0+8
C_0+4	C_0+6	C_0+8	C_0+10
C_0+6	C_0+8	C_0+10	C_0+12

It is also found that the quantisation matrices given in Figure 2-29 (page 42) are not suitable for the transform domain Wyner-Ziv residual codec because of the modified pdf of the DC coefficients. Therefore, based on extensive experimental results, the quantisation matrices illustrated in Figure 4-4 are proposed.

4	4	0	0
4	0	0	0
0	0	0	0
0	0	0	0

(Q1)

8	4	0	0
4	0	0	0
0	0	0	0
0	0	0	0

(Q2)

8	4	4	0
4	4	0	0
4	0	0	0
0	0	0	0

(Q3)

8	8	4	0
8	4	0	0
4	0	0	0
0	0	0	0

(Q4)

16	8	8	0
8	8	0	0
8	0	0	0
0	0	0	0

(Q5)

16	8	8	4
8	8	4	0
8	4	0	0
4	0	0	0

(Q6)

16	8	8	4
8	8	4	4
8	4	4	0
4	4	0	0

(Q7)

32	16	8	4
16	8	4	4
8	4	4	0
4	4	0	0

(Q8)

Figure 4-4: Proposed quantisation matrices

4.1.3 Experimental Results

The technique presented in this section is implemented in the codec discussed in [18]. Therefore, the rate distortion performance of the proposed technique has been compared against the technique presented in [18]. In the simulations, a number of test video sequences have been considered. Figure 4-5, Figure 4-6 and Figure 4-7 illustrate the rate-distortion performance for the first 100 frames of the *Salesman*, *Mother and Daughter* and *Mobile* QCIF video sequences. Rate distortion plots contain the rate and PSNR values of the Wyner-Ziv frames. Other test conditions are as follows:

- Frame rate: 30 fps
- GOP length: 2
- DCT size: 4×4
- The original key frames are considered to be available at the decoder (as in [18]).
- Error detection mode: Ideal error detection (Bit error rate threshold: 10^{-3})

Figure 4-5, Figure 4-6 and Figure 4-7 show that, there is a significant PSNR gain up to 0.6 dB, or up to 50% reduction in the bit rate compared to the technique presented in [18].

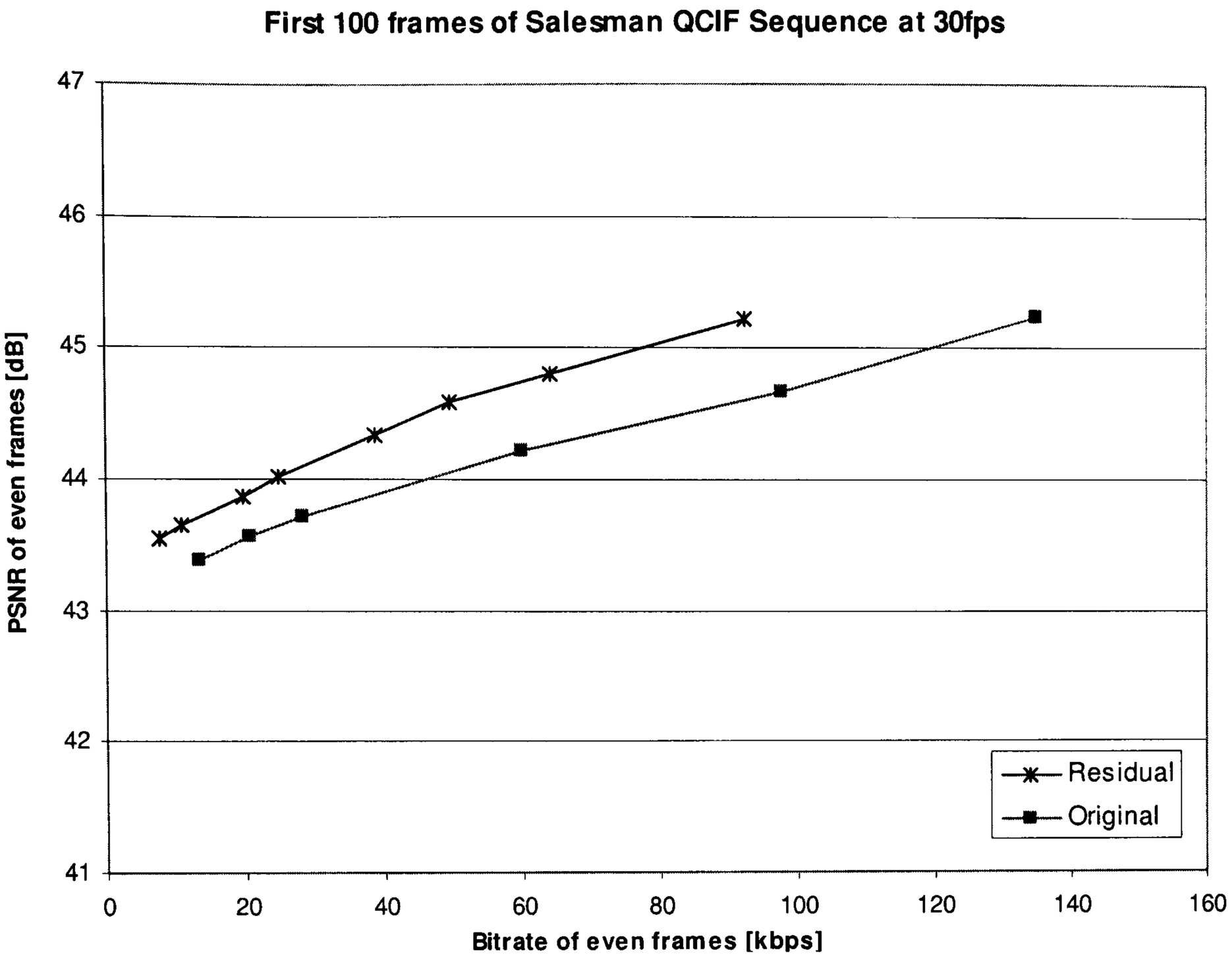


Figure 4-5: RD performance comparison of WZ frames for *Salesman* sequence

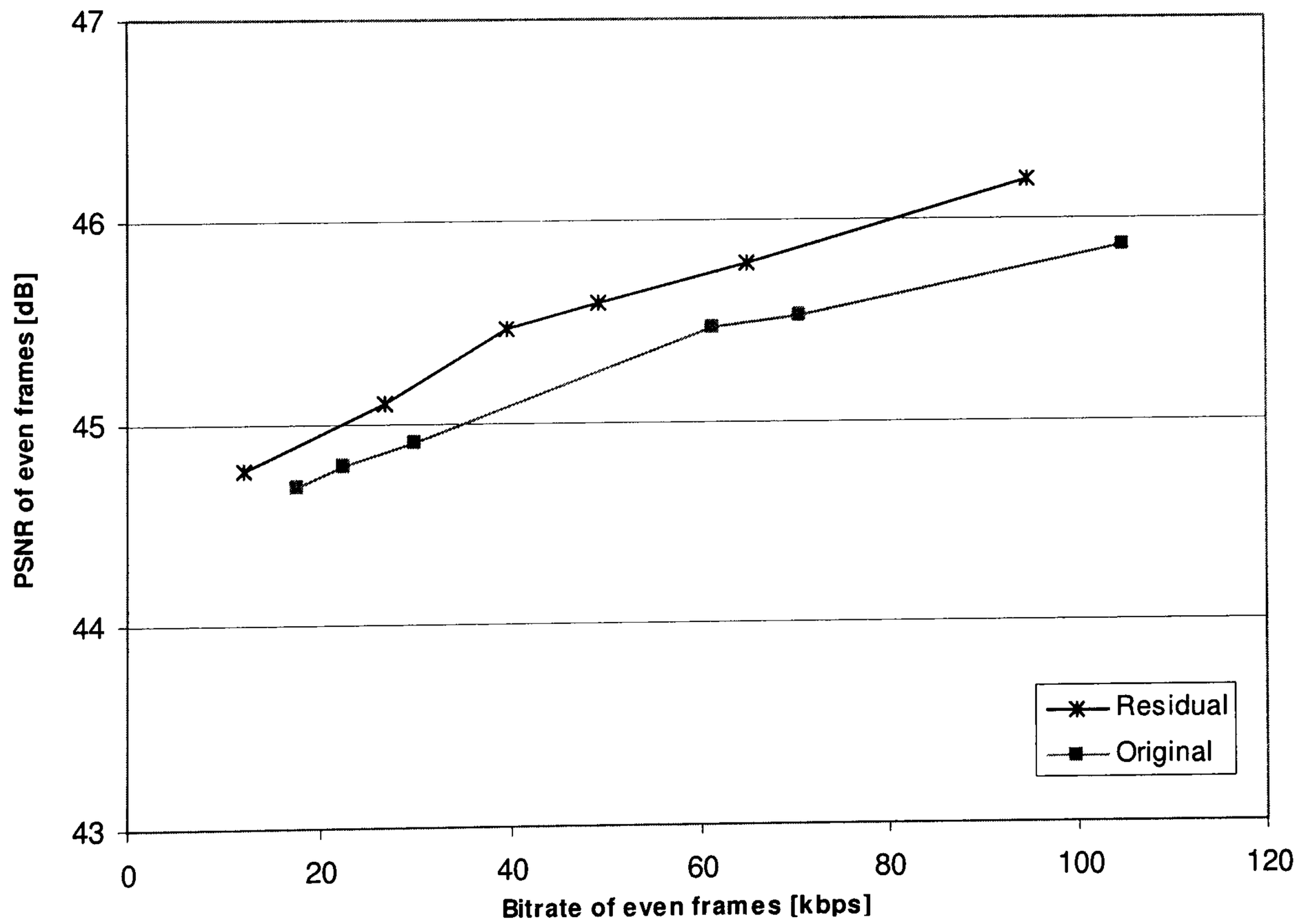


Figure 4-6: RD performance comparison of WZ frames for *Mother and Daughter* sequence

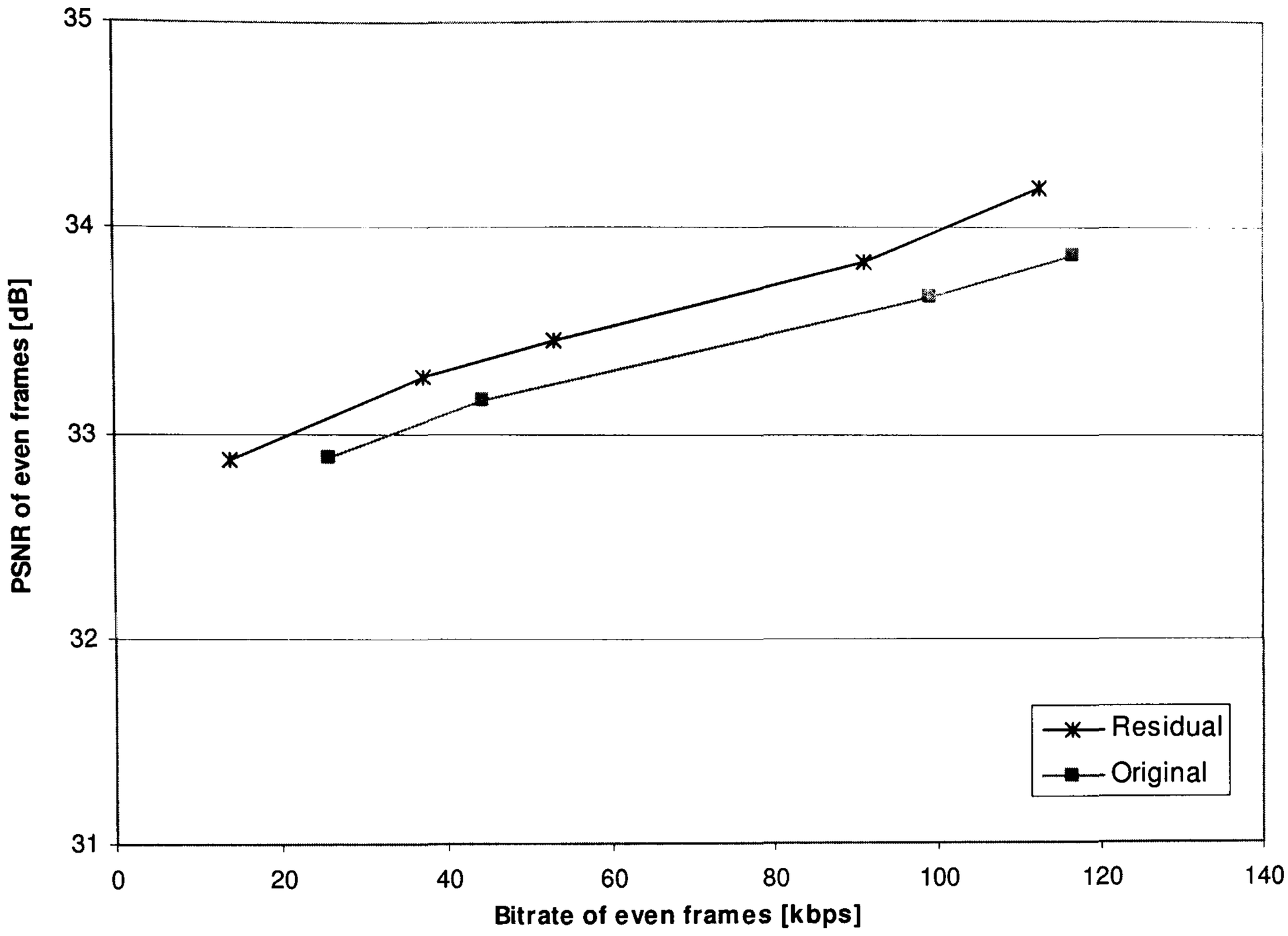


Figure 4-7: RD performance comparison of WZ frames for *Mobile* sequence

4.2 Non-linear Quantisation

The objective of this work is to propose a novel non-linear quantisation algorithm for DVC in contrast to the linear quantiser that has been traditionally used in the common transform domain DVC implementations. It is observed that when the residual image generated by taking the frame difference compared to the preceding reference frame is considered, all the transform coefficients demonstrate a similar probability density function with a contrastingly high density near the zero mean. The vast majority of the DCT coefficients are concentrated in the close vicinity of this peak distribution point. The motivation for the non-linear quantisation approach discussed in this section is to exploit this probability distribution so that small values, which make up a major portion of the input residual signal, are represented with a higher precision so that a better picture quality and therefore improved compression efficiency can be achieved.

4.2.1 Proposed Solution

The proposed quantisation algorithm builds upon the basic hypothesis that each sample element in the input sequence has an unequal contribution to the output video quality depending on the relative magnitude of the element. More specifically, in the case of transform domain DVC, it is assumed that the final decoded picture quality is more sensitive to the small incremental variations of the current frame compared to the reference frame than the large variations, primarily due the very high probability density of the minor variations. The probability density function (pdf) of the DC coefficients of the DCT transformed residual video signal, calculated using the incremental variations of the current Wyner-Ziv frame compared to the reference frame (key frame), is illustrated in Figure 4-2 (page 73). The dotted line depicts the pdf of the DC coefficients (similarly AC coefficients in Figure 4-3 (page 73)) of the original Wyner-Ziv frames, while the solid line represents the same for the residual signal considering the incremental difference with respect to the predecessor key frame, with a GOP size of 2. The contrasting peak in the latter curve shows the concentration of the information content at the very small signal magnitudes in the residual signal. In this section, a non-linear quantisation algorithm is proposed to exploit this concentrated distribution of the information.

In the proposed algorithm, the non-linear quantisation is implemented by passing each DCT coefficient band through a non-linear transformation module incorporated into the quantiser. The reverse process is performed in the reconstruction module at the decoder. The design of the optimum non-linear transformation function for the proposed quantiser is a new research problem. Here, this problem is addressed by evaluating a number of candidate formulas including high

order polynomials and exponential functions as defined in Equations (4.3), (4.4) and (4.5) using a large number of video sequences with different characteristics:

$$y = sign(x)\left(a_3|x|^3 + a_2|x|^2 + a_1|x|\right) \tag{4.3}$$

$$y = a\,sign(x)\left(e^{b.|x|} - 1\right) \tag{4.4}$$

$$y = \beta\,sign(x)\left(1 - e^{-c.|x|/\beta}\right) \tag{4.5}$$

Where, x and y represent the input and output of the non-linear transformation respectively, and a_1, a_2, a_3, a, b, β and c are constants that represent the level of non-linearity. To select the best candidate for the non linear quantisation, we carried out some tests with the above formulas and some of these results (for the *Foreman* QCIF sequence) are shown in Figure 4-8. Results clearly illustrate that the function described in Equation (4.5) shows the best performance. Therefore, we considered this function for all of our simulation results presented in this chapter. Test conditions for this evaluation are:

- Frame rate: 15Hz
- GOP length: 2
- Frames within the sequence: All frames
- $a_1 = 0.2$, $a_2 = 4.10^{-4}$, $a_3 = 8.10^{-7}$
- $a = 185$, $b = 2.10^{-3}$
- $\beta = 500$, $c = 1$

Where a_1, a_2, a_3, a, b, β and c are experimentally optimised.

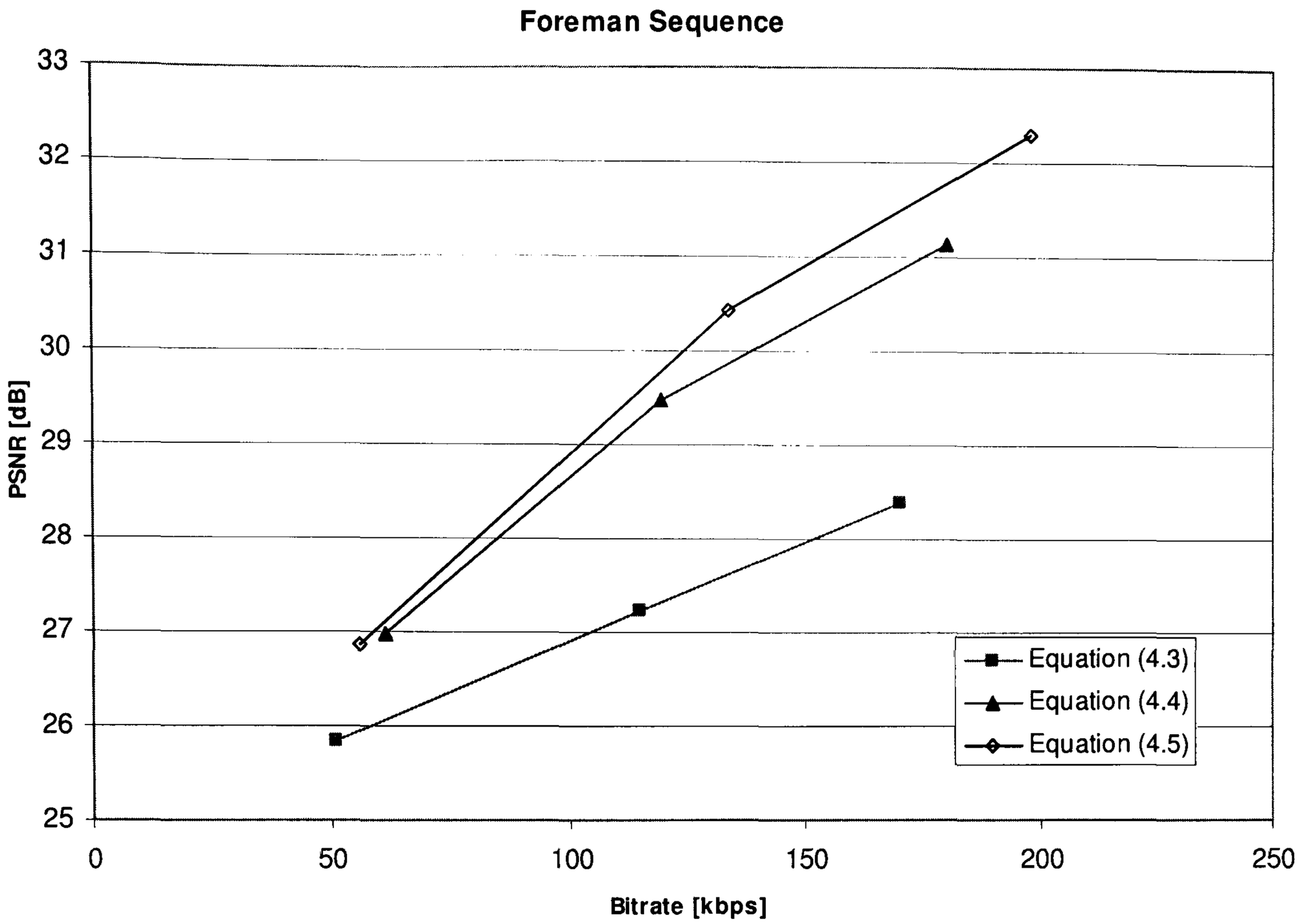


Figure 4-8: Evaluation results for the non-linear transform functions

Figure 4-9 gives a graphical representation of the transform function depicted in Equation (4.5). The variable quantisation bin sizes derived by the non-linear quantisation are illustrated in Figure 4-10, where W represents the quantisation step size with respect to the linear quantiser and W' represents the smallest step size for the non-linear quantiser. The reconstruction function is also modified to incorporate the inverse-non-linear-transformation as defined by:

$$y = -\frac{\beta}{c} sign(x) \ln\left(1 - \frac{|x|}{\beta}\right) \tag{4.6}$$

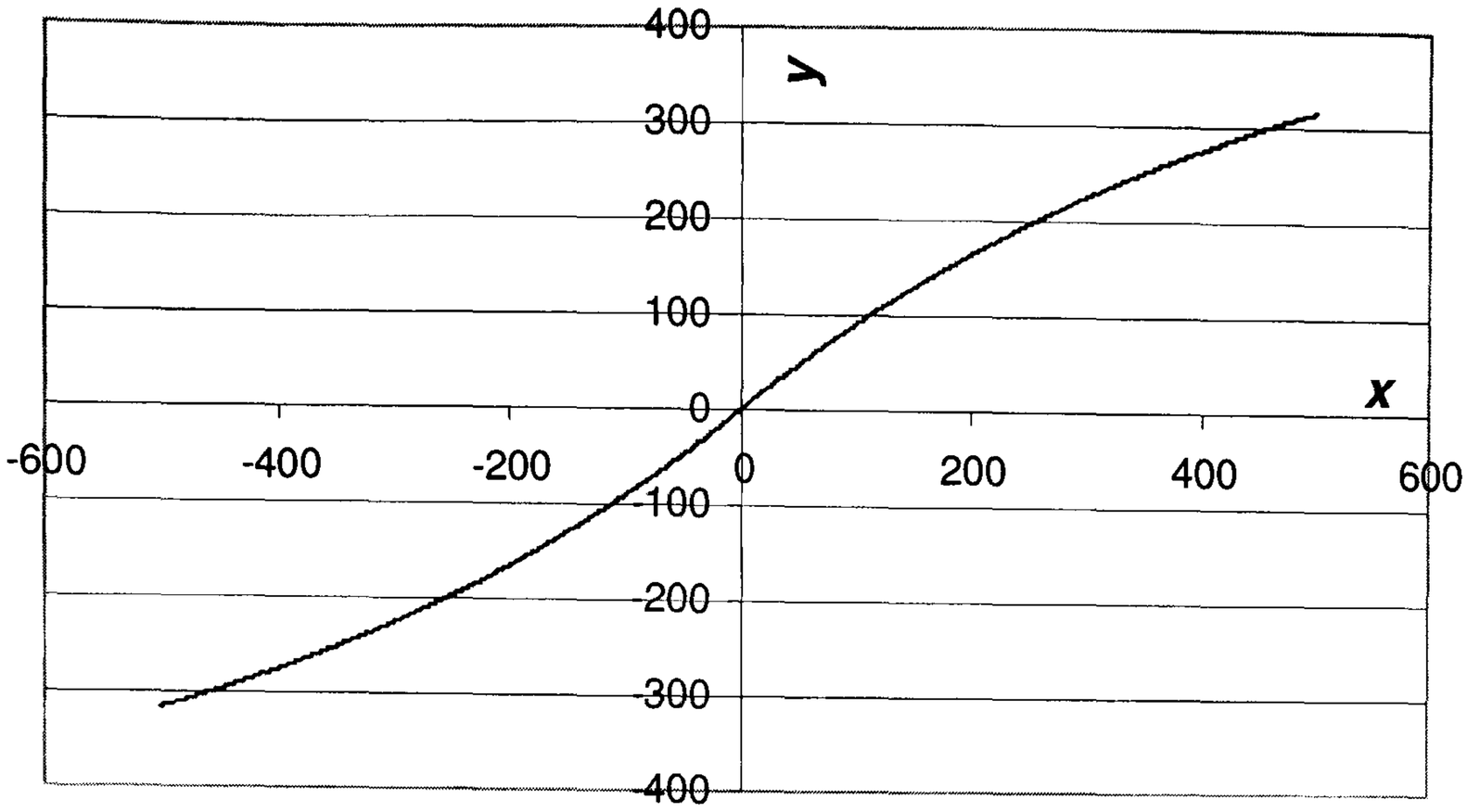


Figure 4-9: Non-linear transform function

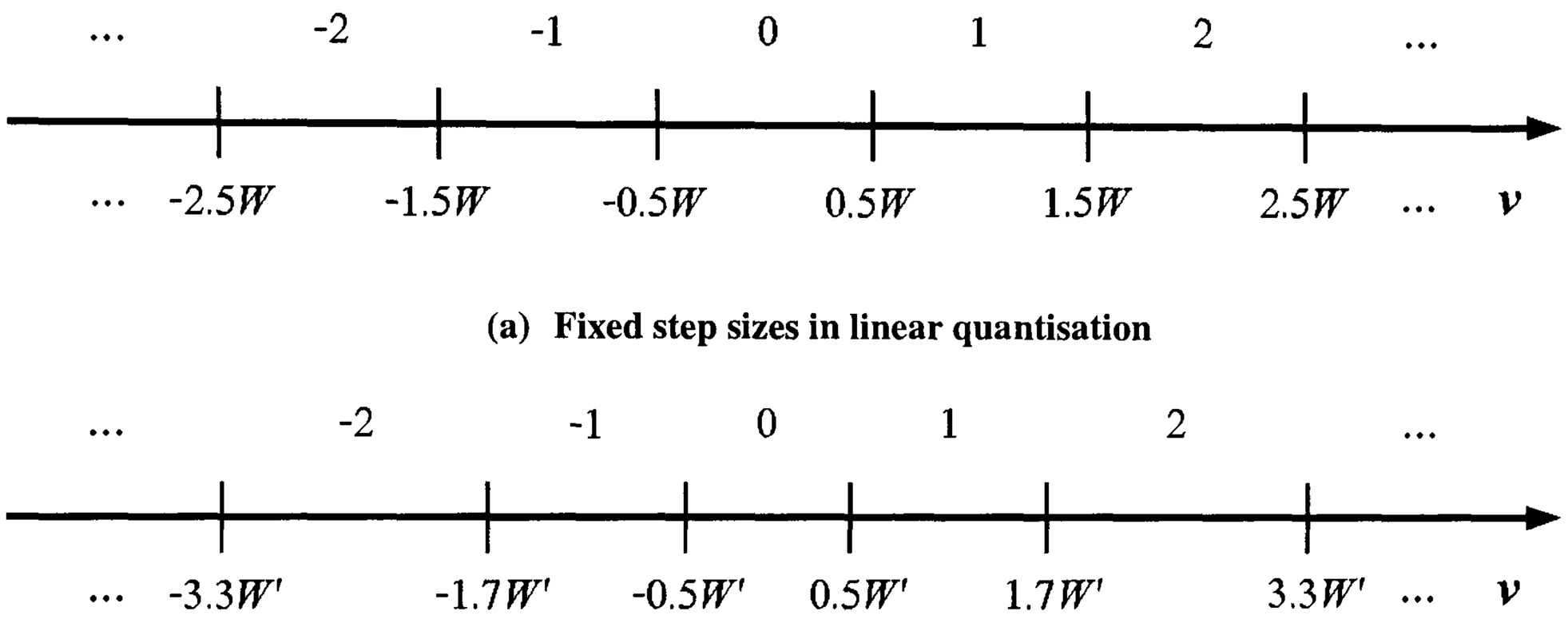


Figure 4-10: Linear and non-linear quantisation step sizes

Figure 4-11 shows the comparison of linear and non-linear quantisers. In this figure, Equation (4.5) is used for the nonlinear quantisation, where W is the step size of the linear quantiser. In this figure, it can be seen that the step sizes and therefore the quantisation error is decreased for the smaller values by using nonlinear quantisation. As the majority of the coefficients are in this range i.e. close to zero, this may help decreasing the overall distortion. However, for larger values the quantisation error is increased. PSNR calculation is linked to the total square of the error as given in Equation (4.7) and the large errors directly influence the distortion. Considering the fact that large errors dominate the overall distortion, optimisation of the non-linearity plays an

important role. In DVC, reconstruction is not only based on the quantised (decoded) values but also the side information. Therefore, quantisation noise affects the output when the predicted and decoded quantised values are different. In this work, the non-linearity is optimised based on empirical results obtained by examining several test sequences (β and c are constants to represent the level of non-linearity).

$$PSNR = 10\log_{10}\left(\frac{255^2}{1/N \sum_{all(i,j)} \left(Y_{ref}(i,j) - Y_{dec}(i,j)\right)^2}\right) \tag{4.7}$$

Where, N is the total number of pixels in the image, $Y_{ref}(i,j)$ and $Y_{dec}(i,j)$ are the pixel values of the reference and decoded images.

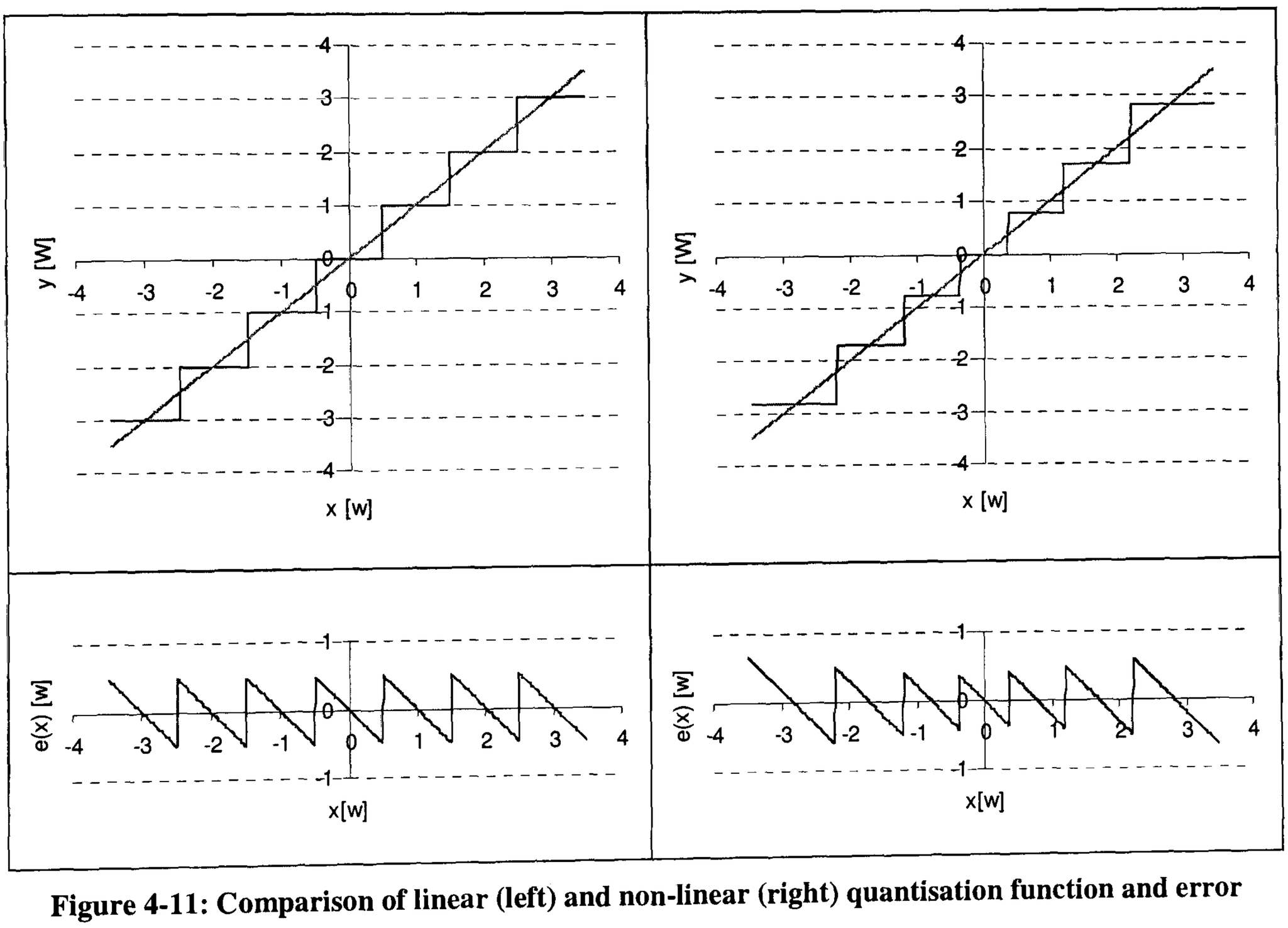


Figure 4-11: Comparison of linear (left) and non-linear (right) quantisation function and error

The proposed quantisation algorithm is incorporated into the existing transform domain DVC framework [18]. The modified codec architecture is depicted in Figure 4-12. First, a frame difference operation is performed to derive the residual frames both at the encoder and the decoder. At the encoder, the difference between the current Wyner-Ziv frame (X_{2i}) and the previous key frame (X_{2i-1}) is taken and the resulting residual frame is 4×4 block-wise DCT transformed. Sub-frames are formed from the sixteen DCT coefficient bands and then the non-linear quantisation is performed for each DCT band sub-frame. At the decoder, side information is obtained by taking the difference of the motion interpolated frame (Y_{2i}) and the previous key frame (X_{2i-1}). In a similar process to that performed at the encoder, the DCT band coefficients are grouped together and then non-linearly quantised. This resulting bit stream is fed into the turbo decoder and the reconstruction blocks. Turbo decoder receives parities in small amounts, and requests for more parity until the bit error rate for the bit plane is lower than a given threshold (10^{-3}). After decoding and reconstruction, the inverse non-linear transform and IDCT operations are performed. Finally, the previous frame (i.e. X_{2i-1}) is added and the reconstructed frame X'_{2i} is obtained.

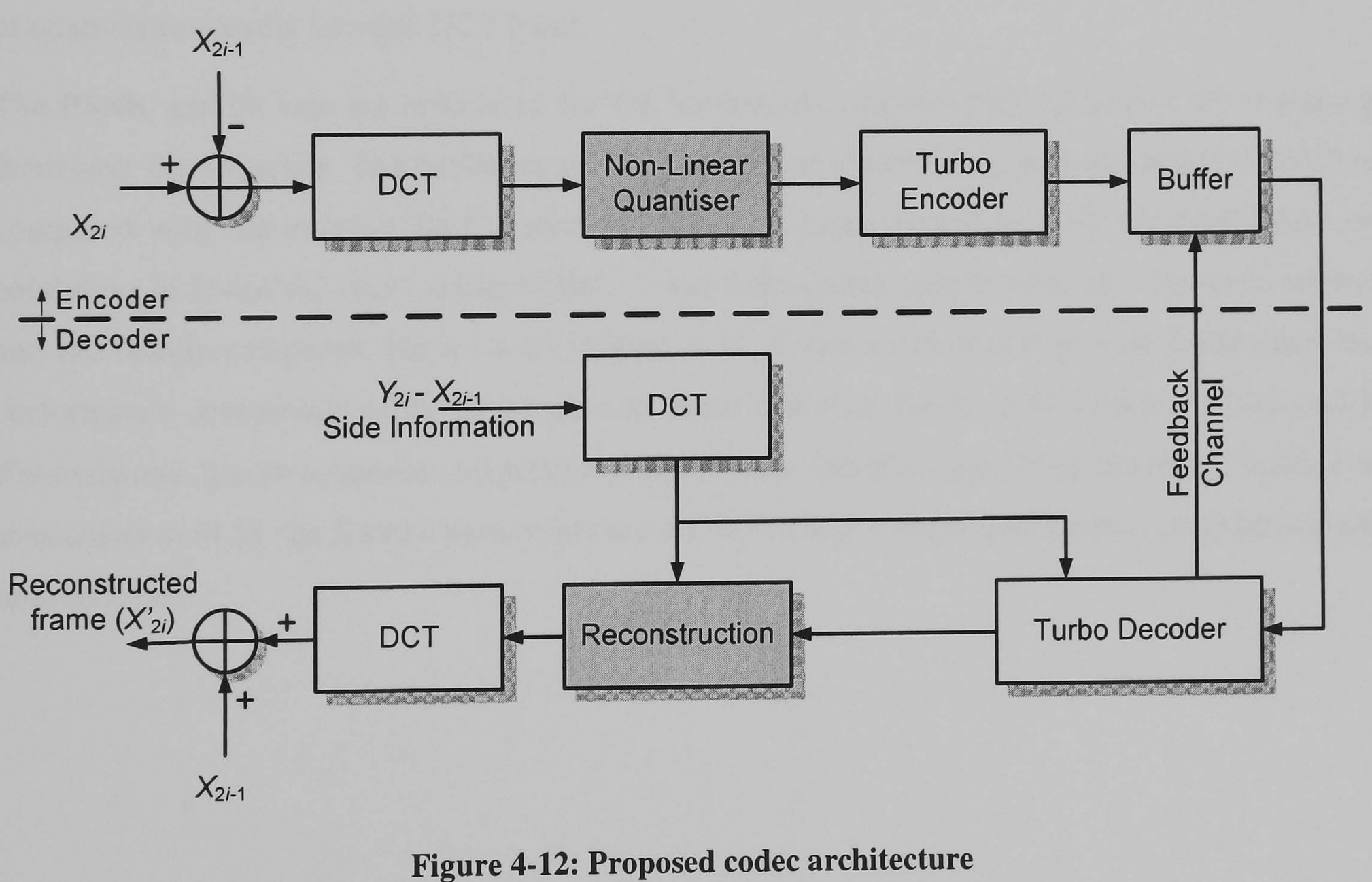


Figure 4-12: Proposed codec architecture

4.2.2 Experimental Results

The RD performance of the modified transform domain DVC codec with the proposed non-linear quantisation is tested using *Hall-Monitor*, *Coastguard*, *Foreman* and *Soccer* QCIF video sequences considering all frames within each sequence.

The following test conditions are used for all simulations presented in this section:

- Frame Rate: 15fps
- GOP length: 2
- Key frames are H.264/AVC intra-coded with same average picture quality for both key frames and Wyner-Ziv frames (please see appendix C)
- Error detection mode: Ideal error detection (Bit error rate threshold: 10^{-3})

These sequences are selected to represent different motion levels; *Hall-Monitor* and *Coastguard* with low to medium motion levels, *Foreman* and *Soccer* with medium to high. The bit rate of Wyner-Ziv frames is varied by independently controlling the granularity of the quantiser, using the quantisation matrices shown in Figure 4-4 (page 75). Note that numbers represent the number of quantisation levels for each DCT band.

The PSNR and bit rate are calculated for the luminance component of all frames by averaging them over the sequence. The performance of the proposed algorithm (non-linear residual DVC) is compared with the existing DVC codec which uses a linear quantiser [18]. Under similar test conditions H.264/AVC inter coding (IBIB…) and intra-coding results have also been considered and the results compared. Figure 4-13, Figure 4-14, Figure 4-15 and Figure 4-16 illustrate the performance comparison of the proposed non-linear quantiser for the *Hall-Monitor, Coastguard, Foreman* and *Soccer* sequences respectively. In the tests, the non-linear transformation function is selected as in (4.5); the β and c parameters are set to 500 and 1 respectively based on experimental optimisations.

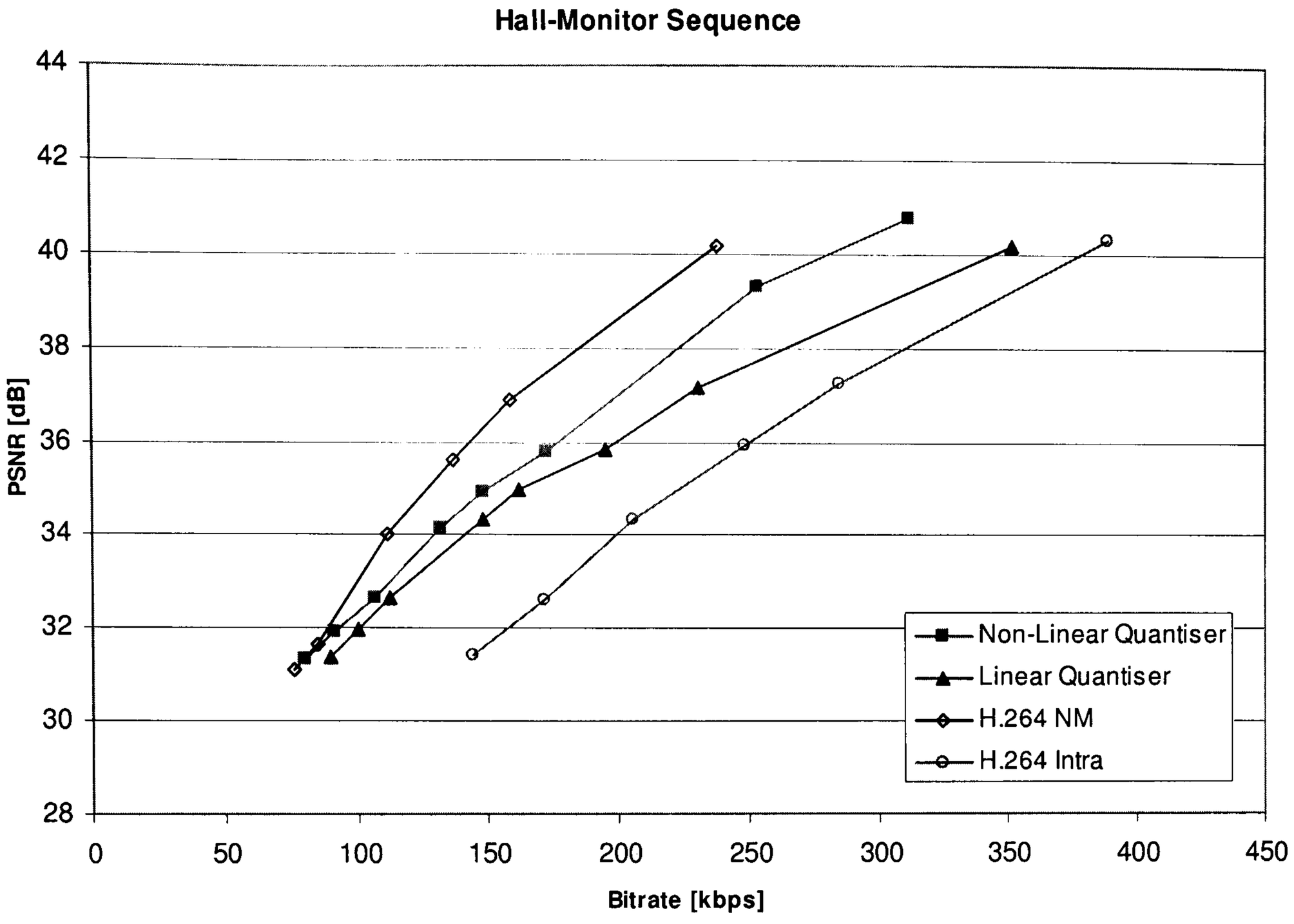


Figure 4-13: Performance comparison for the *Hall-Monitor* test sequence

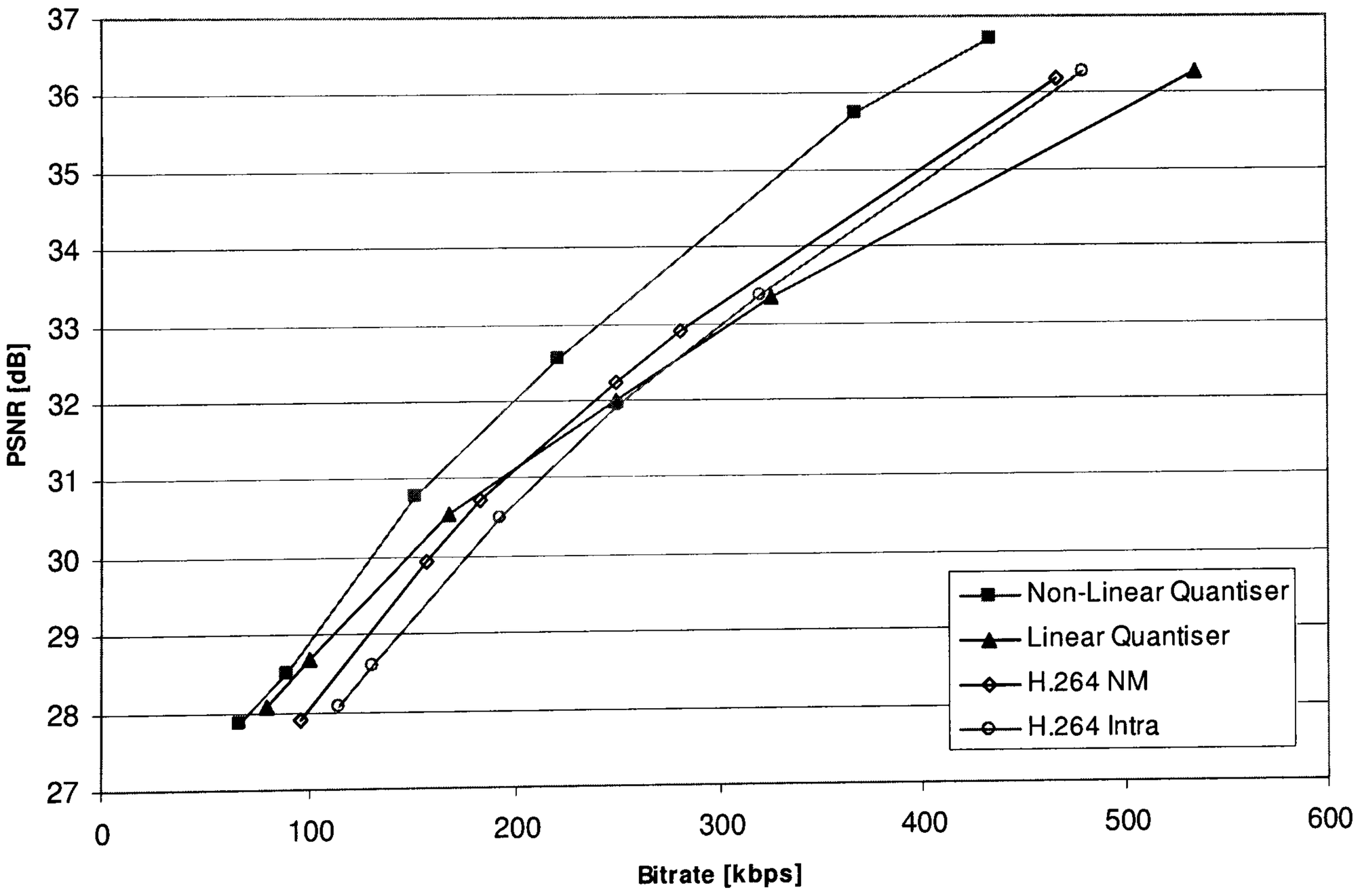


Figure 4-14: Performance comparison for the *Coastguard* test sequence

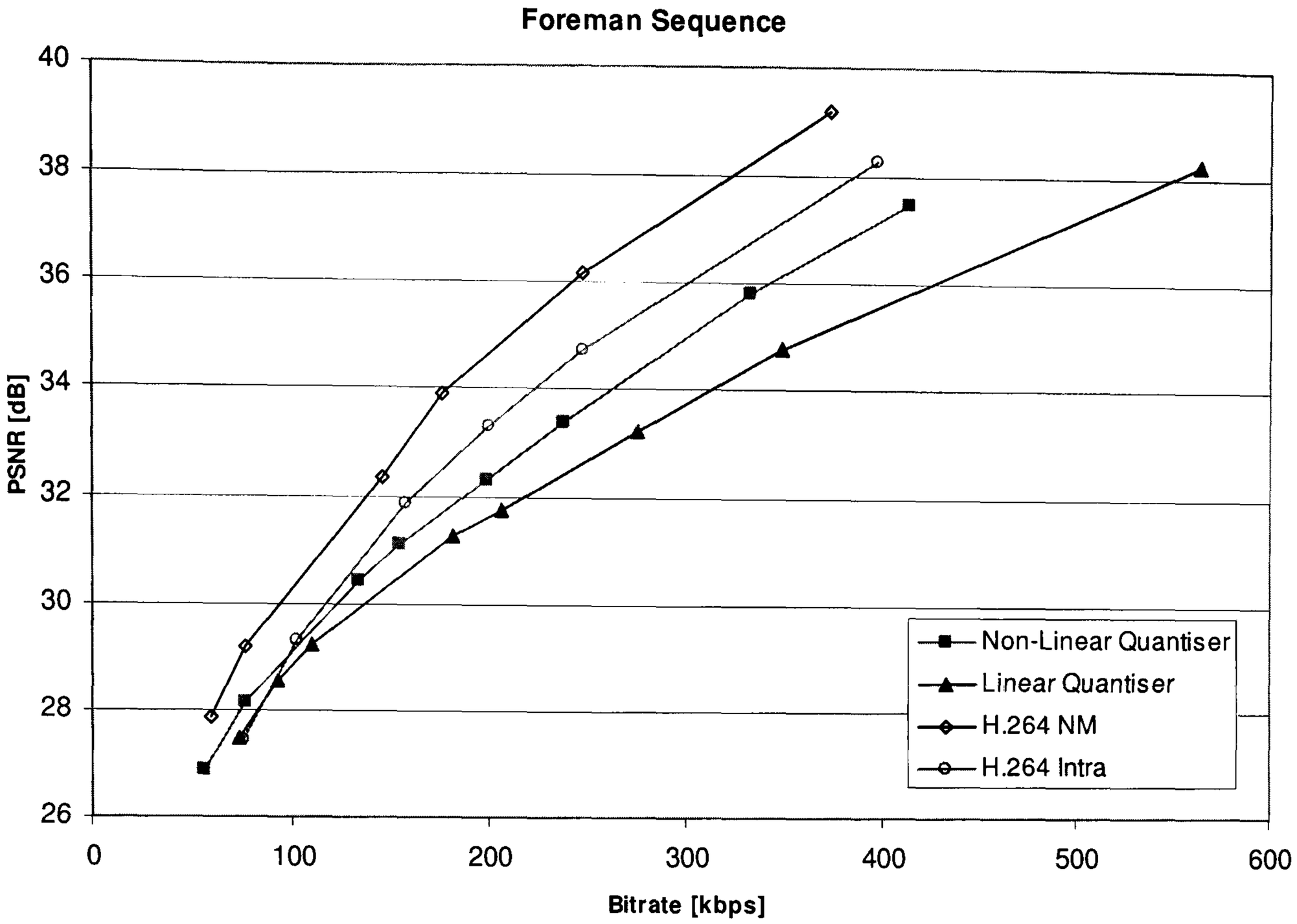


Figure 4-15: Performance comparison for the *Foreman* test sequence

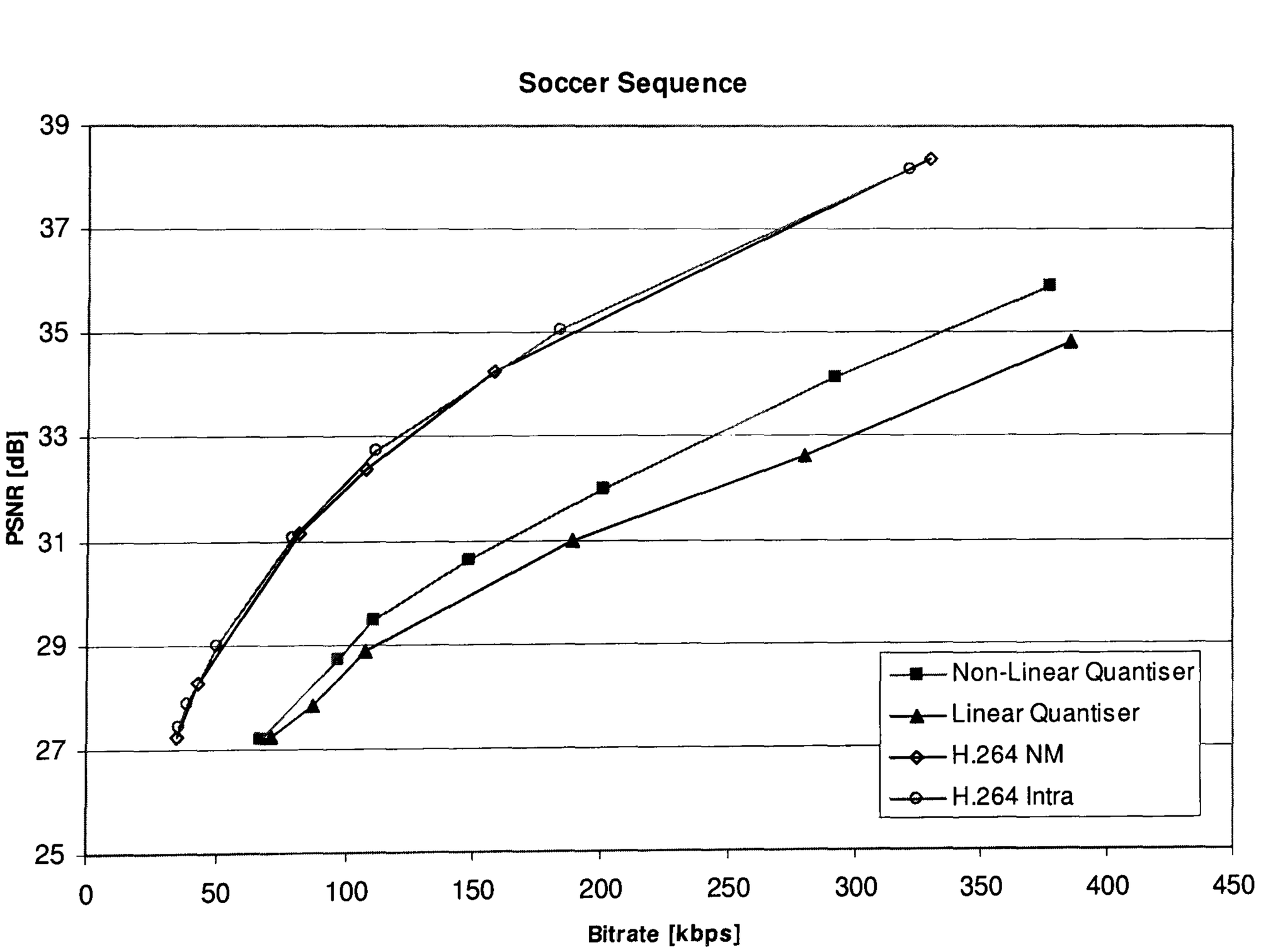
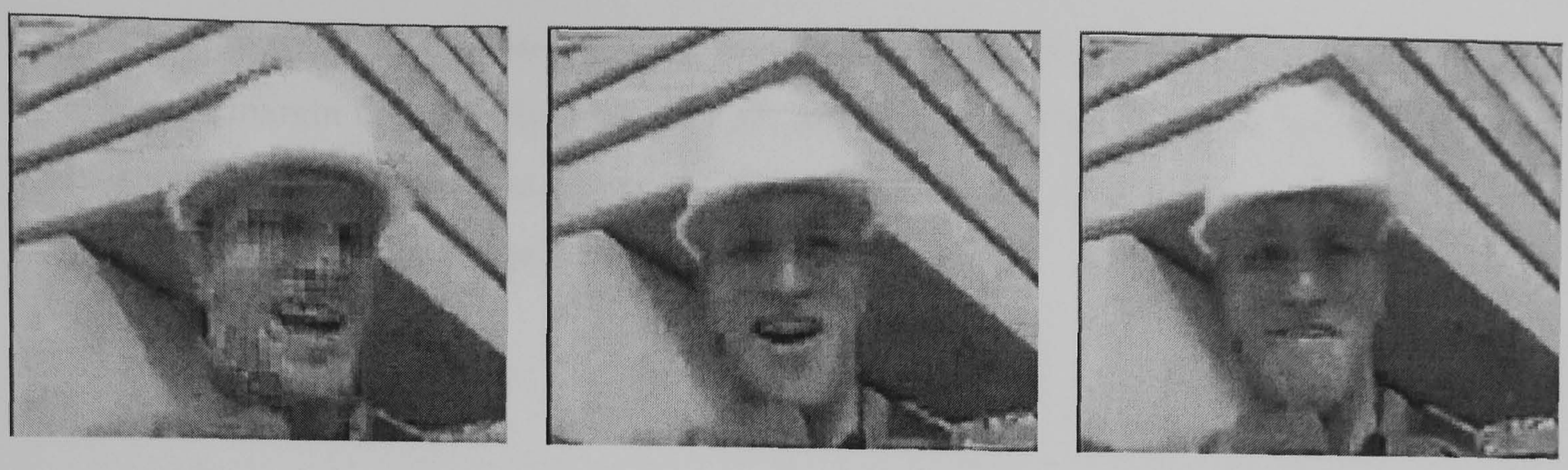


Figure 4-16: Performance comparison for the *Soccer* test sequence



(a)

(b)

Figure 4-17: Images for the *Foreman* test sequence: (a) Linear quantiser and (b) Non-linear quantiser (frame number 8, 16 and 18 from left to right).

Simulation results clearly suggest that the proposed non-linear codec can outperform the DVC codec based on linear quantisation by a significant RD performance margin. Furthermore, results show that the proposed non-linear quantisation can close the gap between the H.264/AVC and DVC significantly. For the *Coastguard* sequence the proposed solution outperforms H.264/AVC inter-coding. The evident performance gain can be attributed to the decreased quantisation error in the proposed solution when the coefficients are concentrated around zero and the step sizes are smaller in that region.

It is noted that when there is a significant motion in the video sequence, the side information generated in DVC based on motion interpolation is grossly inaccurate. As a result, the rate distortion performance is notably lower compared to H.264/AVC for which the current frame is available for motion estimation. This is evident in Figure 4-15 (*Foreman*) and Figure 4-16 (*Soccer*).

Figure 4-17 shows frames numbers 8, 16 and 18 of the *Foreman* sequence respectively for the proposed non-linear quantisation based DVC codec and the linear quantisation based DVC codec at a constant bit rate of 0.14 average bits per pixel. It can be observed that the proposed solution produces better picture quality.

The results show that the proposed algorithm helps to improve the performance of the DVC codec by a notable margin consistently over different bit rates and different motion levels in the input video sequence. Therefore, the proposed modification to the quantisation process will move forward DVC research quite significantly.

4.3 Conclusion

In this chapter, a novel residual coding technique and a non-linear quantisation technique are proposed and implemented into the transform domain DVC codec.

In the first section, a new technique for transform domain DVC is proposed. In this new scheme, the encoder exploits both spatial and temporal redundancies by DCT based coding and taking the pixel-wise difference of two consecutive frames. Subsequently, residual frames are coded as Wyner-Ziv frames. Moreover, a novel quantisation technique is proposed to optimise the rate-distortion performance. Simulation results show that there is an improvement in PSNR by up to 0.6 dB compared to reference transform domain Wyner-Ziv codec [18], at the expense of a minor increase in computational complexity.

In the second part of this chapter, a nonlinear quantisation technique is presented for considering the signal characteristics of residual frames. The DCT coefficient bands (DC and AC coefficients) of the residual signal of the Wyner-Ziv frames, derived by taking the difference with the preceding key frame, demonstrate a distinct pdf with a very high probability concentration around the near zero mean. Therefore, a non-linear quantisation algorithm is proposed for the DVC codec to exploit the dominating contribution from the relatively small values of the DCT coefficient bands concentrated near zero as described above. The simulation results depict a consistently improved RD performance at all bit rates when different test video sequences with varying motion levels are considered. The proposed non-linear quantisation technique can improve the performance by up to 2 dB.

The residual coding and non-linear quantisation techniques in this chapter have been published in conferences or journals (please see appendix B).

Chapter 5

5 Unidirectional Distributed Video Coding

DVC based video codecs proposed in the literature generally include a reverse (feedback) channel between the encoder and the decoder. This channel is used to communicate the dynamic parity bit request messages from the decoder to the encoder resulting in an optimum dynamic variable rate control implementation. However it is observed that this dynamic feedback mechanism is a practical hindrance in a number of practical consumer electronics applications.

In this chapter, a novel transform domain Unidirectional Distributed Video Codec (UDVC) without a feedback channel is proposed. A simple encoder rate control algorithm is used in order to eliminate the feedback channel of the existing DVC codecs while keeping the encoder complexity as low as possible. First, all Wyner-Ziv frames are divided into macro blocks. A simple metric is used for each block to represent the correlations between the corresponding blocks in the adjacent key frame and the Wyner-Ziv frame. Based on the value of this metric, parity is allocated dynamically for each block. These parities are either stored for offline processing or transmitted to the DVC decoder for online processing.

With UDVC, the number of parity bits received can be less or more compared to DVC with a feedback channel. When the parities are more than required, the resultant quality is the same but the total bit rate is higher. However, the quality of the output frames is lower when the received parity bits are not sufficient.

In the second part of this chapter, an improved reconstruction algorithm is utilised in order to improve the quality of the reconstructed frame. The improved reconstruction technique is implemented on the UDVC codec which will be described in the following section.

The remainder of this section explores the proposal of a Unidirectional DVC codec, whilst the implementation of an improved reconstruction algorithm is dealt with in section 5.2 and finally section 5.3 concludes this chapter.

5.1 Feedback Channel in DVC

One of the foremost difficulties in DVC, which has been overlooked so far, is the close coupling of the encoder with the decoder through a feedback channel. Over this feedback channel, the decoder drives the encoder dynamically, optimising the amount of information needed for decoding the video frame successfully [1]. Unfortunately, this decoder driven architecture is a practical hindrance in a number of applied scenarios, particularly where continuous data storage for offline processing is demanded and when a reverse communication channel is not available. Consequently, DVC is not viable for a range of consumer applications where a light weight encoder design is of particular importance. Therefore, to broaden the DVC usability it is necessary to overhaul the current DVC architecture, specifically, the dependency of the feedback channel. The outcome of such a radical structural change is the Unidirectional DVC (UDVC) architecture. This will eradicate the major practical difficulty in deploying DVC in consumer electronics applications such as mobile multimedia messaging applications, disposable camcorders, wireless PC cameras, network camcorders, etc as shown in Figure 5-1.

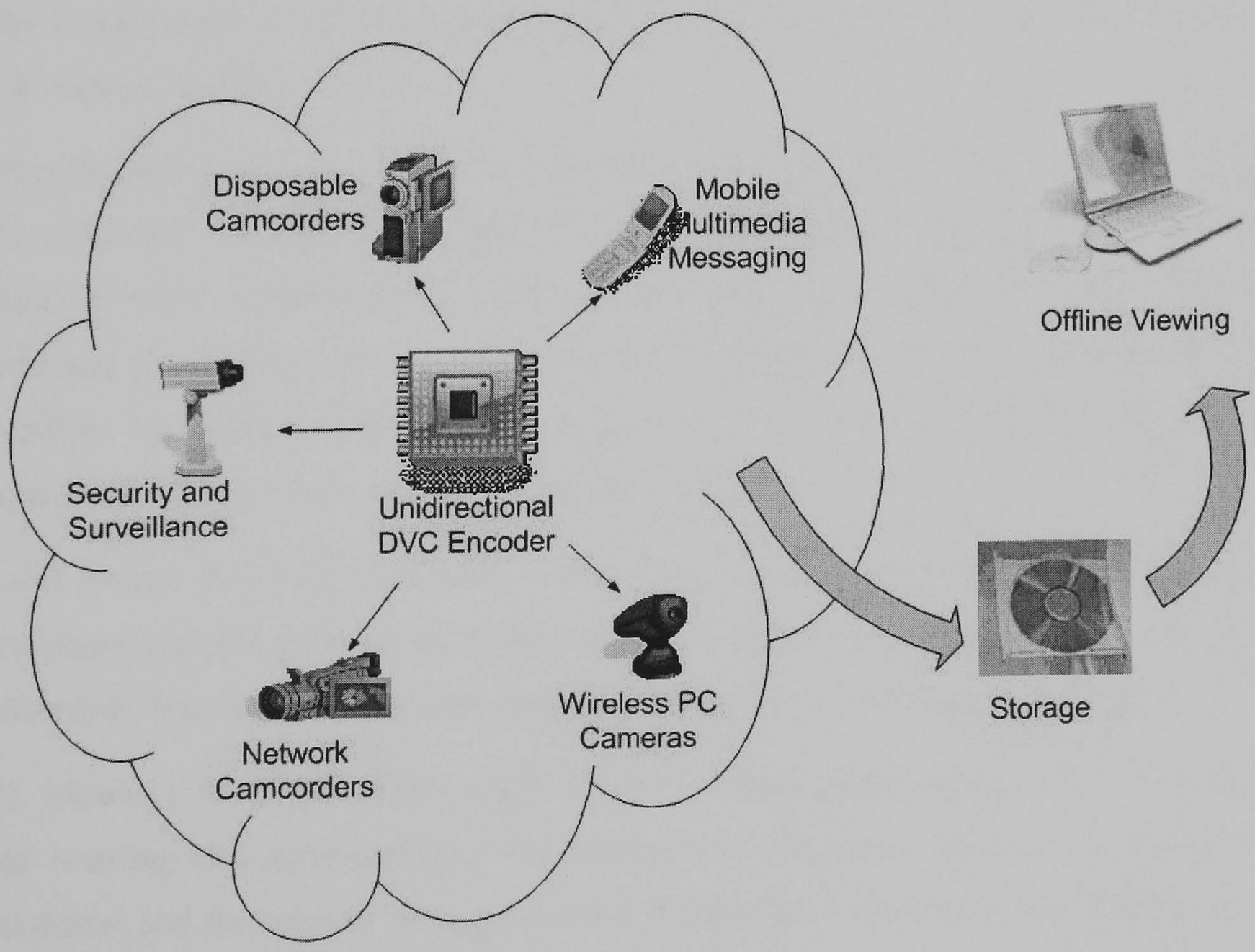


Figure 5-1: Application examples of UDVC

Most of the research in DVC is to improve the rate distortion characteristics keeping the closely coupled codec architecture. However, recently, there have been several attempts to develop a unidirectional DVC architecture eliminating the need of a feedback channel. Artigas and Torres suppressed the feedback channel by estimating the quality of the side information, which is usually predicted at the decoder, at the encoder [75]. The quality estimation is based on the difference between the original frame and the average of the previous and the next key frames. Then the bit rate is determined using an algorithm based on empirical results obtained by testing several standard test sequences.

Adikari et al. proposed a unidirectional DVC technique using parallel Wyner-Ziv coders [76]. The two parallel Wyner-Ziv encoders generate two parity bit streams using independent bit streams. Therefore, the effects of the localised motion is scattered within the frame by using interleavers. Encoder rate is defined by the user based on the quality, the output and the system resources. Weerakkody et al. further improved this algorithm by employing side information refinement [77]. The proposed pixel domain DVC technique heavily relies on iterative decoding with gradual refinement of side information using spatial and temporal predictions at the decoder end to achieve an acceptable decoded picture quality. However, this technique employs ideal error detection method to detect the erroneous pixels. In addition, these multiple predictions and decoding cycles incur huge computational complexity and therefore it has severe limitations in practical implementations.

Morbee et al. proposed a rate allocation technique for pixel domain DVC in [78]. The proposed algorithm estimates the adequate number of bits for every frame. The optimum rate is obtained by estimating the error probabilities for each bit plane. Moreover the residual error probabilities are estimated and if it is above a threshold parity bits are ignored. This work was improved by the same authors by estimating the error probabilities for each pixel and adapting the proposed technique for the case of lossy key frame coding in [79].

Brites and Pereira proposed an encoder rate control algorithm for transform domain DVC [80]. They estimated the side information at the encoder using fast motion estimation to facilitate parity rate estimation. These operations have increased the encoder complexity significantly.

In [81], Martinez et al. proposed a half feedback based pixel domain DVC technique using machine learning and determining if the feedback is necessary. The correlation between the residual frame and the number of requests over the feedback channel is exploited to get a lower complexity encoder. Later in [82], Martinez et al. improved this idea and proposed a feedback free DVC architecture using machine learning.

5.2 Proposed Unidirectional Solution

Considering the unavailability of the feedback channel for many practical cases, a novel transform domain unidirectional DVC codec is proposed in this section. The increase in encoder complexity is kept low by using simple operations such as frame difference; unlike the codec presented in [80] which utilises fast motion estimation at the encoder.

The proposed codec architecture is depicted in Figure 5-2. The encoder is quite similar to the DVC encoder discussed in the previous sections. However, a unidirectional architecture is considered in the transform domain.

First, DCT is applied for each $b \times b$ block. $b \times b$ DCT band frames are formed from this block-wise DCT transformed frame as shown in Figure 5-3. Each of these frames represents a unique DCT band. These frames are divided into $m \times n$ rectangular blocks, which are used as the basic unit for the motion activity based dynamic parity allocation. Following sections describe this algorithm.

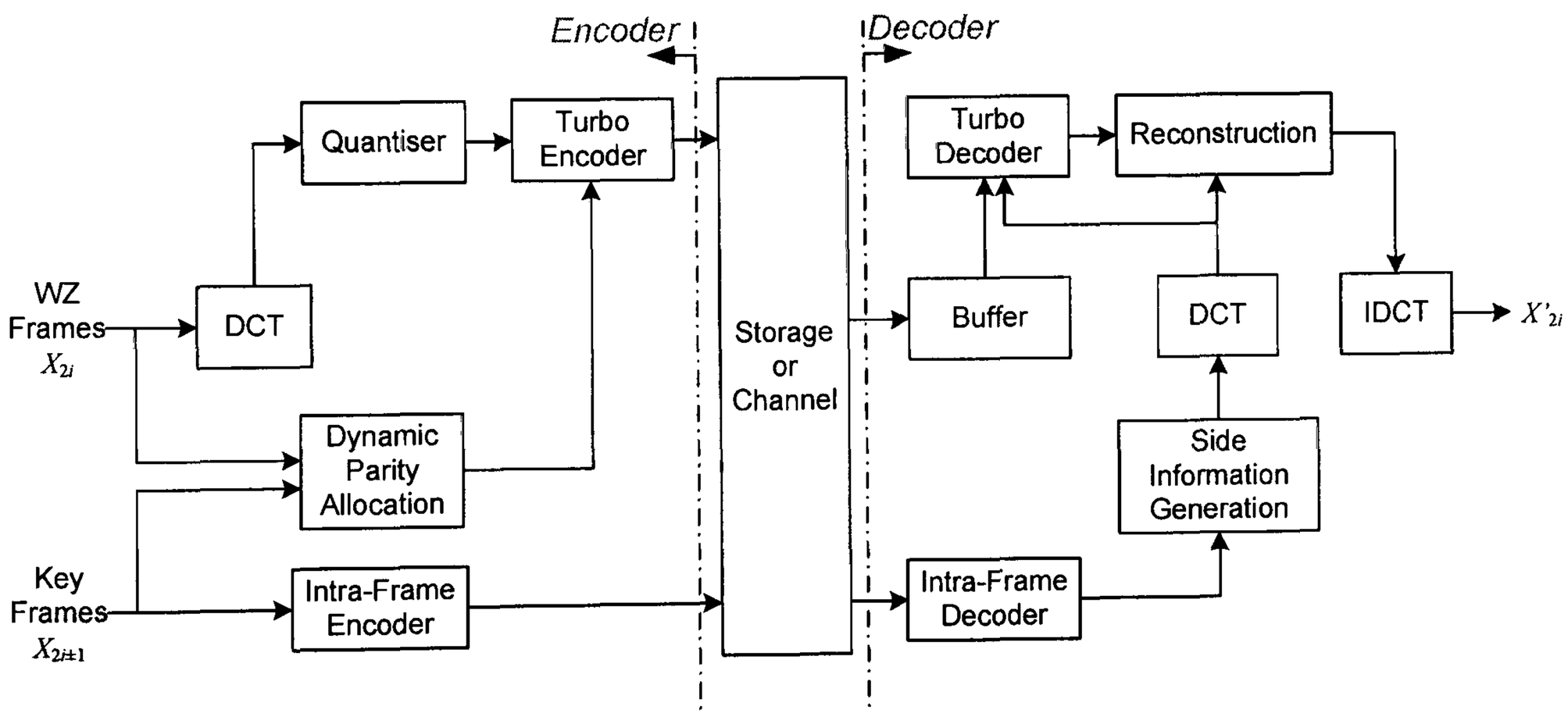


Figure 5-2: Proposed codec architecture

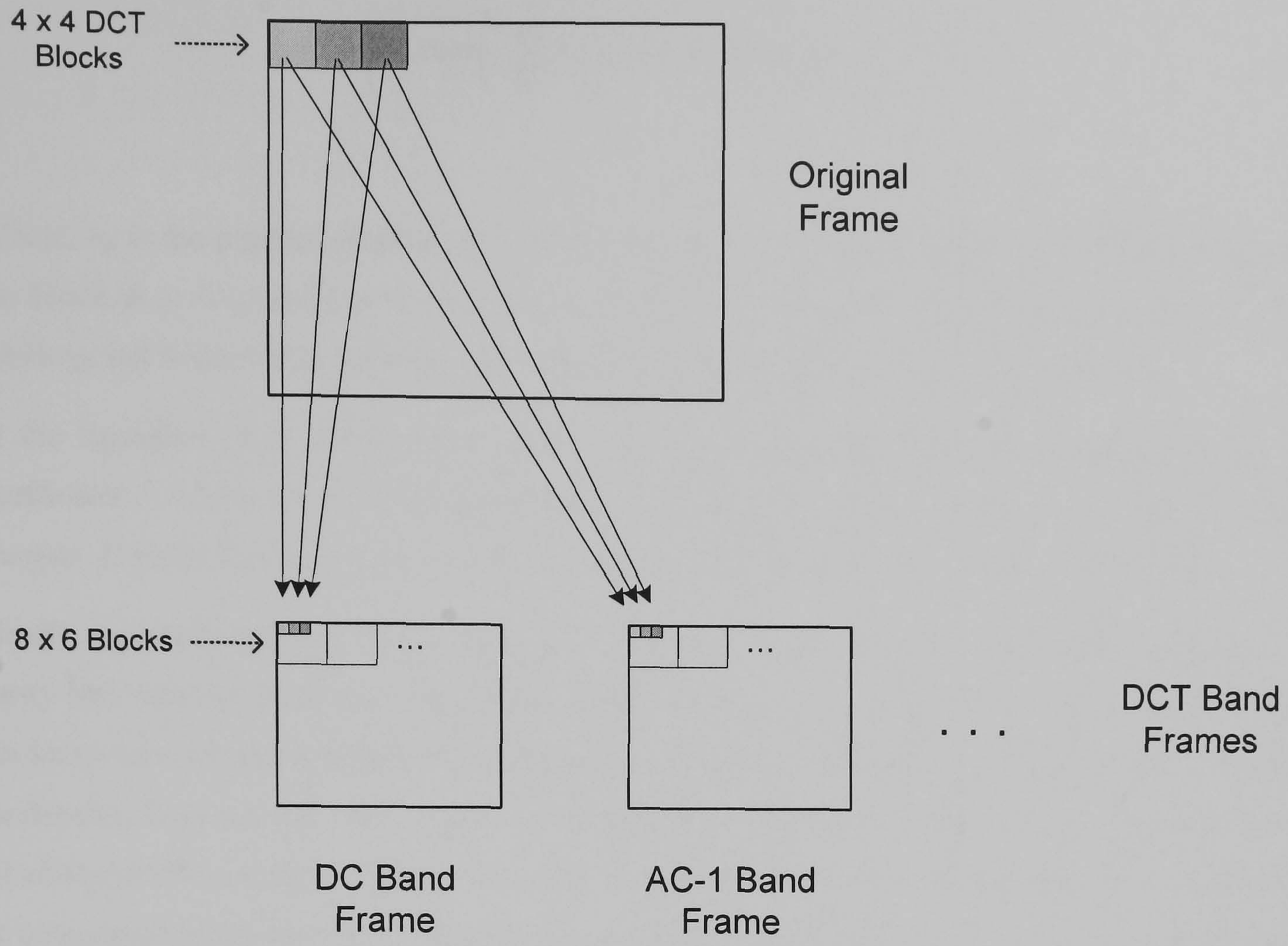


Figure 5-3: Blocks architecture

5.2.1 Dynamic Parity Allocation

For an input data block of length $m \times n$, the turbo encoder generates $2mn$ number of parity bits in the two constituent Recursive Systemetic Convolutional (RSC) encoders [77]. This parity bit stream should be the compressed output of the video encoder; hence the full parity bit sequence can not be moved forward due to the large content. Therefore, a periodic puncturing function, which selects bits from the parity stream in a periodic pattern, is applied for composing the final video encoder output. The corresponding bits are generally selected symmetrically from both parity streams from the two constituent RSC encoders in the turbo encoder. Since the remaining bits are dropped, there could be a significant information loss in this process depending on the implementation algorithm. Thus, puncturing process could be identified as the only trade-off of video quality and compression ratio since there is no feedback path. This puncturing is based on the motion activity between corresponding blocks between the key frame and the Wyner-Ziv frame.

At the encoder Wyner-Ziv frames and key frames are compared, the difference between them is calculated to estimate the motion activity. The average absolute difference between a given block of the Wyner-Ziv frame and the co-located block of the previous key frame is calculated and parity is allocated based on the motion activity as shown in the following equation:

$$n_B = \min\left(M, \text{floor}\left(c\sum_B |X_{2i}(x,y) - X_{2i-1}(x,y)|\right)\right) \tag{5.1}$$

Where, n_B is the number of parity bits that is sent to the decoder for each DCT band bit plane of the block B, c is a constant to control the bit rate, X_{2i} is the Wyner-Ziv frame and X_{2i-1} is the previous key frame. M is the maximum number of parities generated for a given block.

In the Equation (5.1), the residual signal is summed over each block and multiplied by the coefficient c, which needs to be determined by the user according to the bit rate capacity of the channel. If multiplication of the sum and c is larger than M, n_B is equal to M (all parities).

Finally, n_B parity bits are transmitted to the decoder for each block. Information on number of parity bits used for corresponding blocks is also needed to be transmitted since the decoder does not know how much parity are transmitted for each block. Since there is no closed loop feedback mechanism between the encoder and the decoder for the proposed codec, such a dynamic parity bit allocation is essential to control the video quality at the decoder. Furthermore, this process can be considered as an automated supportive mechanism for minimising the bit rate, i.e. maximising the compression.

5.2.2 Channel Update

In DVC, it is assumed that the side information generated by exploiting temporal correlations of the video sequence is a version of the original Wyner-Ziv frame used at the encoder, subjected to noise [1]. It has been reported in literature that the probability distribution of this noise component resembles either Gaussian or Laplacian distributions depending on the statistics of the frame [83]. In this solution, a Laplacian distribution is assumed. Study into the effects of the dynamic variations of the frame statistics on this parameter and thus, to the performance of the codec is a matter of further research and is beyond the scope within this work. In the turbo decoder, each soft-input soft-output (SISO) decoder is implemented using iterative Maximum A Posteriori (MAP) algorithm [84] which employs a hypothetical channel model for systematic bits. As mentioned before, here the assumption is that this channel can be modelled using a Laplacian distribution model. After decoding each DCT band bit plane, this channel model is updated based on the decoded information as in [18]. Previously decoded information is considered while estimating the a posteriori probabilities of the data bits. This update process plays a very important role in the proposed UDVC codec since the decoder needs to depend only on the received parities.

5.2.3 Simulation Results

The rate distortion performance of the proposed transform domain UDVC codec is tested for *Foreman*, *Hall-Monitor* and *Salesman* QCIF (176 × 144) test sequences. These sequences are selected to represent different motion levels; *Foreman* with medium to high; *Hall-Monitor* and *Salesman* with low to medium motion. First 101 frames from each sequence are used for the simulation. The bit rate is varied by independently controlling the granularity of the quantiser, using the quantisation matrices given in Figure 2-29 (page 42) and quantisation parameters for key frames are selected accordingly to match the average PSNR values of key frames and WZ frames. Only the luminance component is considered. PSNR and bit rate are averaged over the sequence.

The proposed UDVC codec is implemented into the DVC codec used in [18], and the results are compared against the same codec. Note that same blocking architecture is considered for DVC with feedback as in proposed UDVC technique.

Other test conditions:

- Frame rate: 30 fps
- GOP length: 2
- DCT size: 4×4
- $m \times n$ is selected as 6×8.
- For the proposed technique c is selected as $\frac{1}{32 \times 24}$.
- For DVC, error detection mode: Ideal error detection (Bit error rate threshold: 10^{-3})

Figure 5-4, Figure 5-5 and Figure 5-6 illustrate the performance comparison of the proposed codec for the *Foreman*, *Hall-Monitor* and *Salesman* sequences respectively. The PSNR of the proposed UDVC codec is lower than the DVC codec with a feedback channel in Figure 5-4. This is acceptable since optimum parity allocation has been considered in the DVC codec with a feedback channel. However, it is observed that this dynamic feedback mechanism is a practical hindrance in a number of practical scenarios, particularly where continuous on-site data storage for offline processing is demanded and also when a reverse communication channel is not available or having one is highly expensive. However, in Figure 5-6, the proposed UDVC codec performs slightly better than DVC with feedback. This is mainly because when the bit rate is lower than the ideal bit rate (bit rate of DVC with feedback) the side information accuracy has a stronger effect on the output. It should be noted that, this sequence (*Salesman*) contains virtually no motion hence the side information accuracy is very high. Corresponding rate distortion points

show the drop in both the bit rate and quality but the overall performance of UDVC is higher in this case.

Figure 5-7 shows frame numbers 12, 50 and 90 of *Foreman* sequence respectively for the DVC with feedback and the proposed UDVC codec. Results are obtained with conditions: QM = 8 for WZ frames and QP = 26 for key frames, which is the last point in the Figure 5-4 for both techniques. It is obvious that the drop in the objective quality is reflected as annoying artefacts in some frames. Therefore, following section (Section 5.3) concentrates on improving the quality of the output of the proposed UDVC technique.

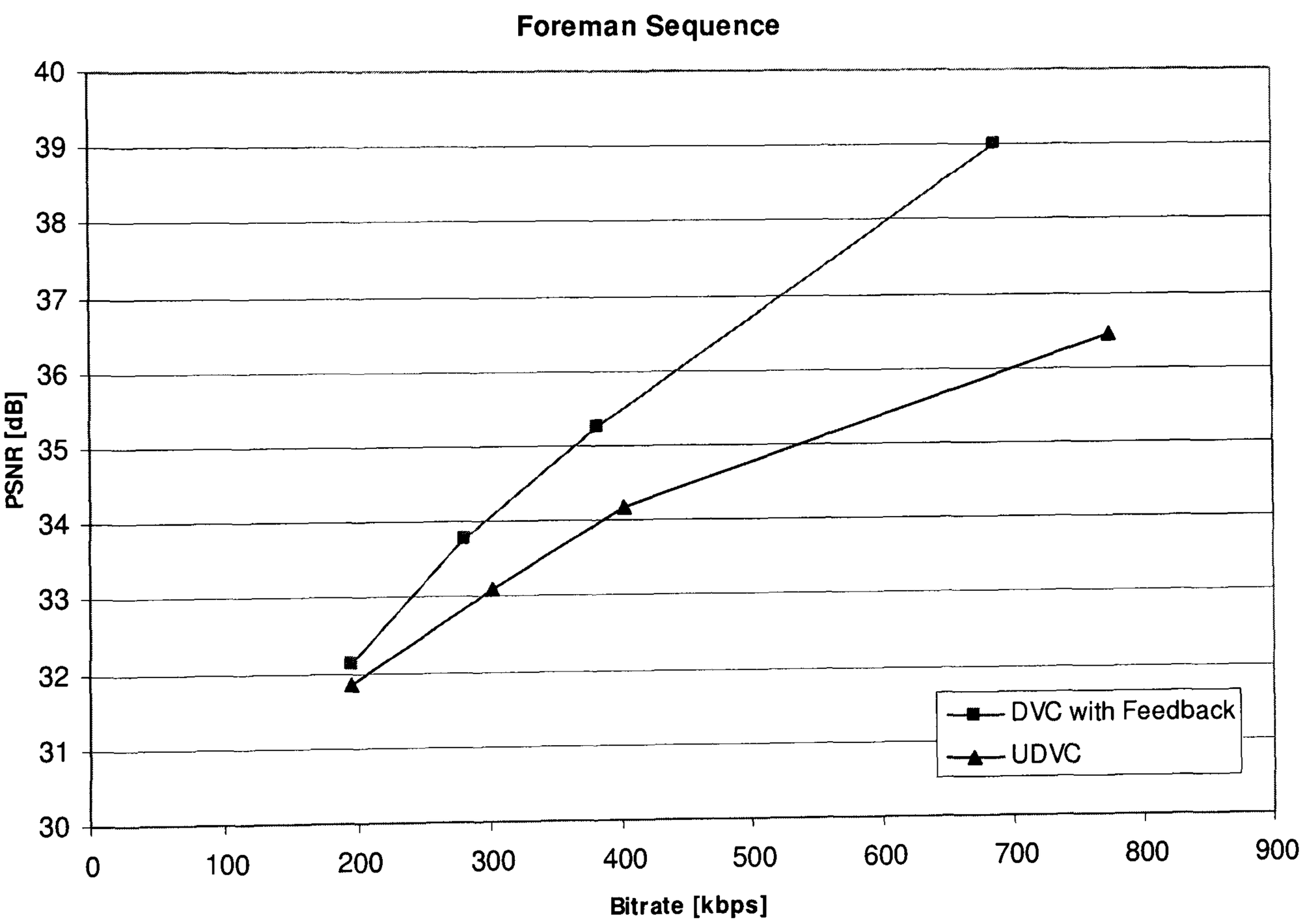


Figure 5-4: Performance comparison for *Foreman* sequence

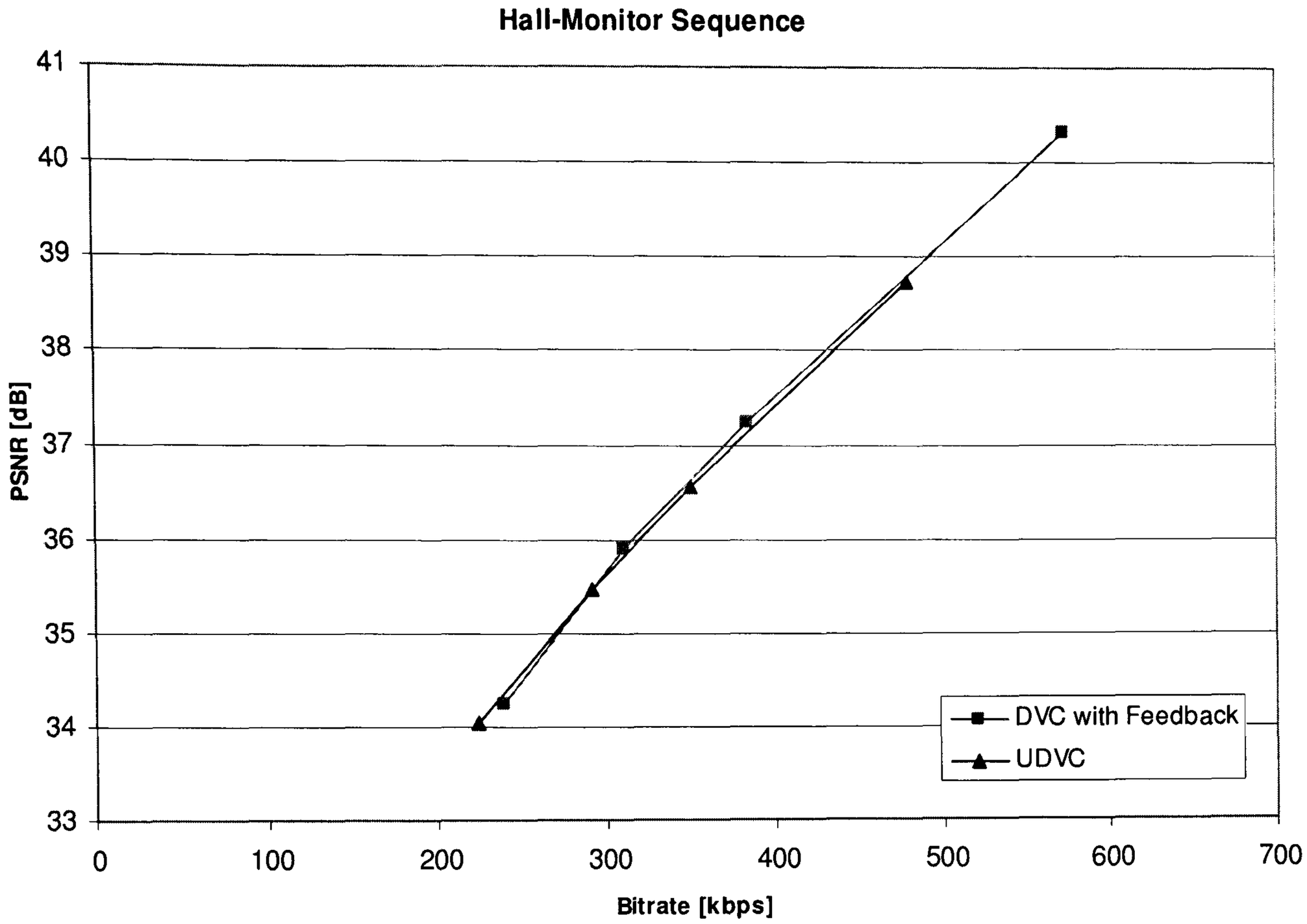


Figure 5-5: Performance comparison for *Hall-Monitor* sequence

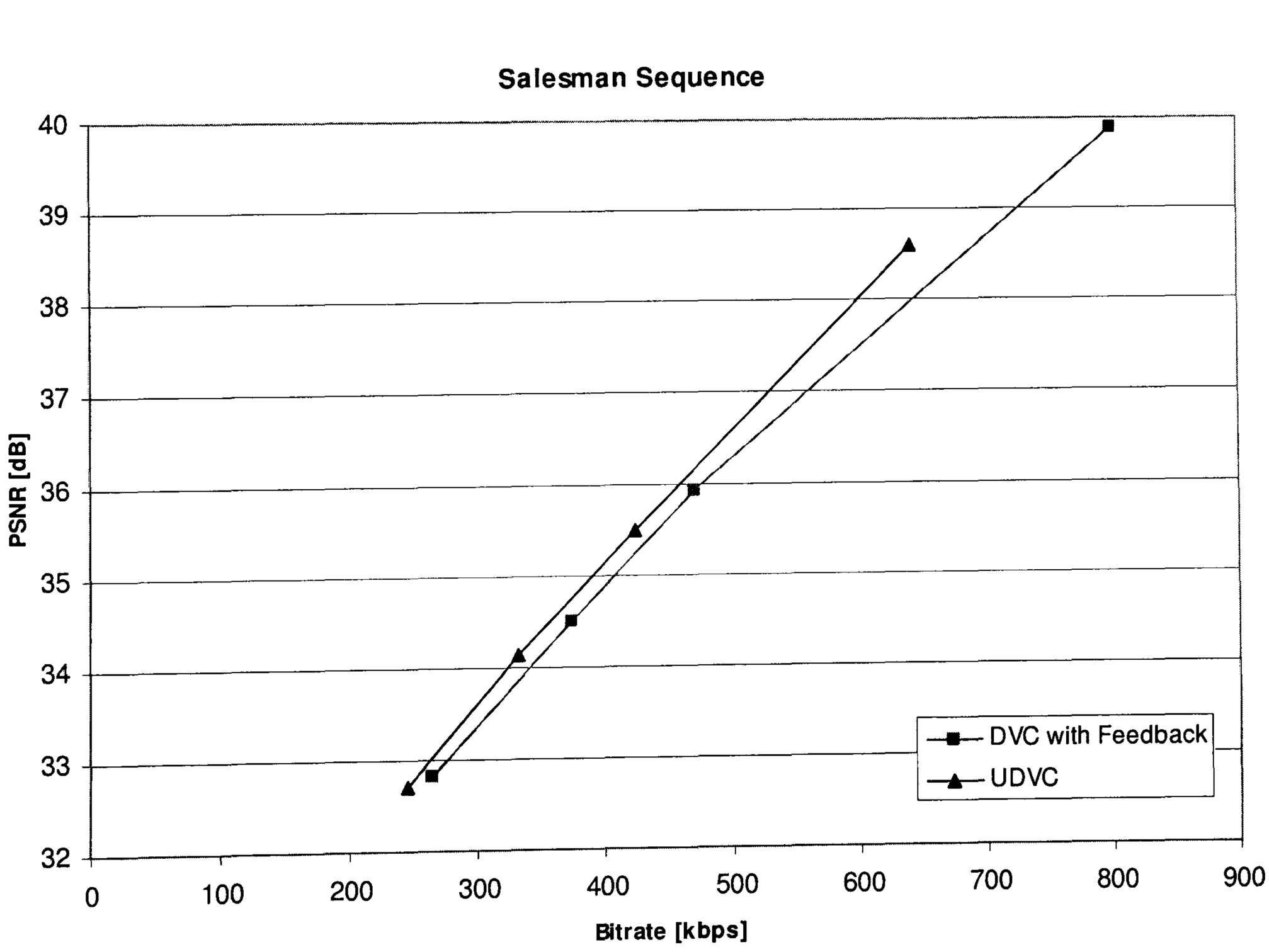
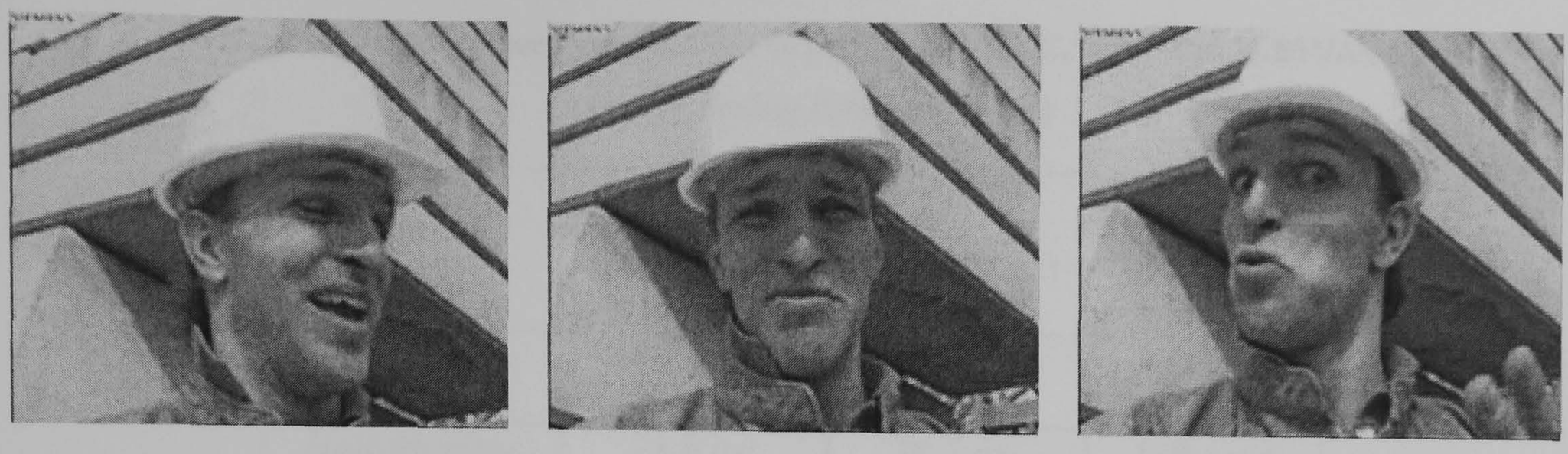
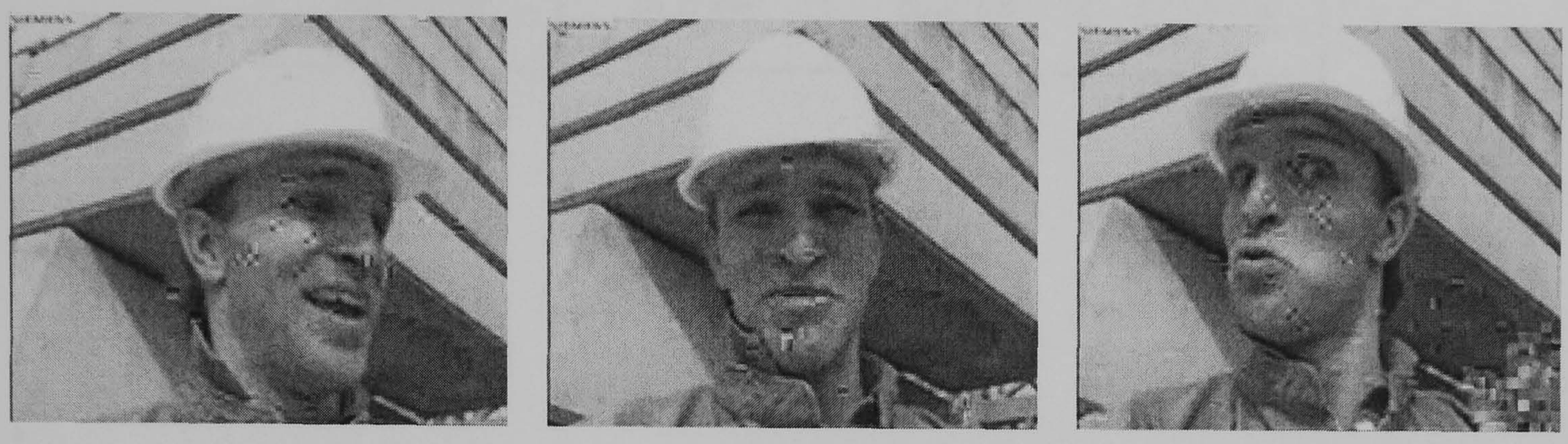


Figure 5-6: Performance comparison for *Salesman* sequence



(a)



(b)

Figure 5-7: Images for the *Foreman* test sequence: (a) DVC with feedback and (b) UDVC (frame number 12, 50 and 90 from left to right).

5.2.4 Computational Complexity

Compared to the DVC with feedback, we have used only one additional block at the encoder to exploit the correlations between the key frame and the Wyner-Ziv frame. However, this is not computationally expensive as it can be seen from Equation (5.1). Table 5.1 gives the encoder speeds of proposed UDVC, DVC and H.264/AVC codec. Performance of the H.264/AVC codec is obtained by restricting motion vectors to zero in bidirectional coded pictures in JM12.4 reference software. Same test conditions are considered for UDVC and DVC as given in the previous section (QM = 4). Key frames/Intra-coded frames are the same for all codecs and only Wyner-Ziv frame (for UDVC and DVC) and B frame (for H.264/AVC) coding speeds are given. Simulations are performed under the same conditions on the same computer. Table 5.1, shows that the encoder complexities of proposed UDVC and DVC are similar and the proposed UDVC is still significantly simple comparing to H.264/AVC.

Table 5.1: Encoder speed comparison of proposed UDVC, DVC and H.264/AVC (Average coding time per frame)

Sequence	UDVC (WZ frames)	DVC (WZ frames)	H.264/AVC (B frames)
Foreman	20.2 ms	18.8 ms	580.5 ms
Hall-Monitor	21.6 ms	20.3 ms	548.4 ms
Salesman	21.6 ms	20.4 ms	573.8 ms

5.3 Improved Reconstruction for the Proposed UDVC Solution

The reconstruction method has been one of the most important processes in DVC. At the encoder, quantisation is performed in order to control the bit rate (or quality) by limiting the number of bits per pixel. Reconstruction, performed at the decoder, is the reverse process of quantisation. One of the widely used reconstruction algorithms is proposed by Aaron et al. in [73]. This technique does not consider any statistical information obtained from the neighbouring pixel for reconstruction. Also, the bit error rate (BER) is not considered in this technique. Later, in [85] and [86], an enhanced reconstruction algorithm is proposed which predicts the reconstructed value using a Laplacian model of the residual frame between the original frame and the side information.

Considering unidirectional DVC schemes, Weerakkody et al. proposed another reconstruction algorithm which makes use of the BER of each bit plane [57]. This technique is adopted for the unidirectional DVC codec proposed in the previous section (Section 5.2) and explained in the following subsection.

5.3.1 Improved Reconstruction Algorithm

Weerakkody et al. modified the reconstruction algorithm based on the assumption that the importance of the side information in the reconstruction process increases when the bit error rates of each decoded bit plane increase. Therefore, for the case when the decoded quantised information is not the same as the side information, the reconstruction output is modelled using the function defined in the following equation [57];

$$\hat{x} = \begin{cases} z^- + \tan(k)(y - z^-) & , y < z^- \\ z^+ + \tan(k)(y - z^+) & , y > z^+ \end{cases} \tag{5.2}$$

Where $\hat{x}$ is the reconstructed pixel, y is the side information, z^- and z^+ are the lower and upper boundaries of the quantisation bin. k is defined as the *recon angle* which is an index showing the noise distribution of the decoded bit stream and the side information.

For the other case, when the decoded quantised pixel is same as the side information, reconstructed value is equal to the side information:

$$\hat{x} = y \quad , \qquad z^- < y < z^+ \tag{5.3}$$

An illustration of the Weerakkody's model is depicted in Figure 5-8. In this figure, q_i represents the quantised information. The *recon angle* can be clearly seen from the gradient of the reconstruction lines. The value of the optimised angle significantly varies depending on the decoded information and the side information. In the transform domain implementation, recon angles are separately optimised for each band [57].

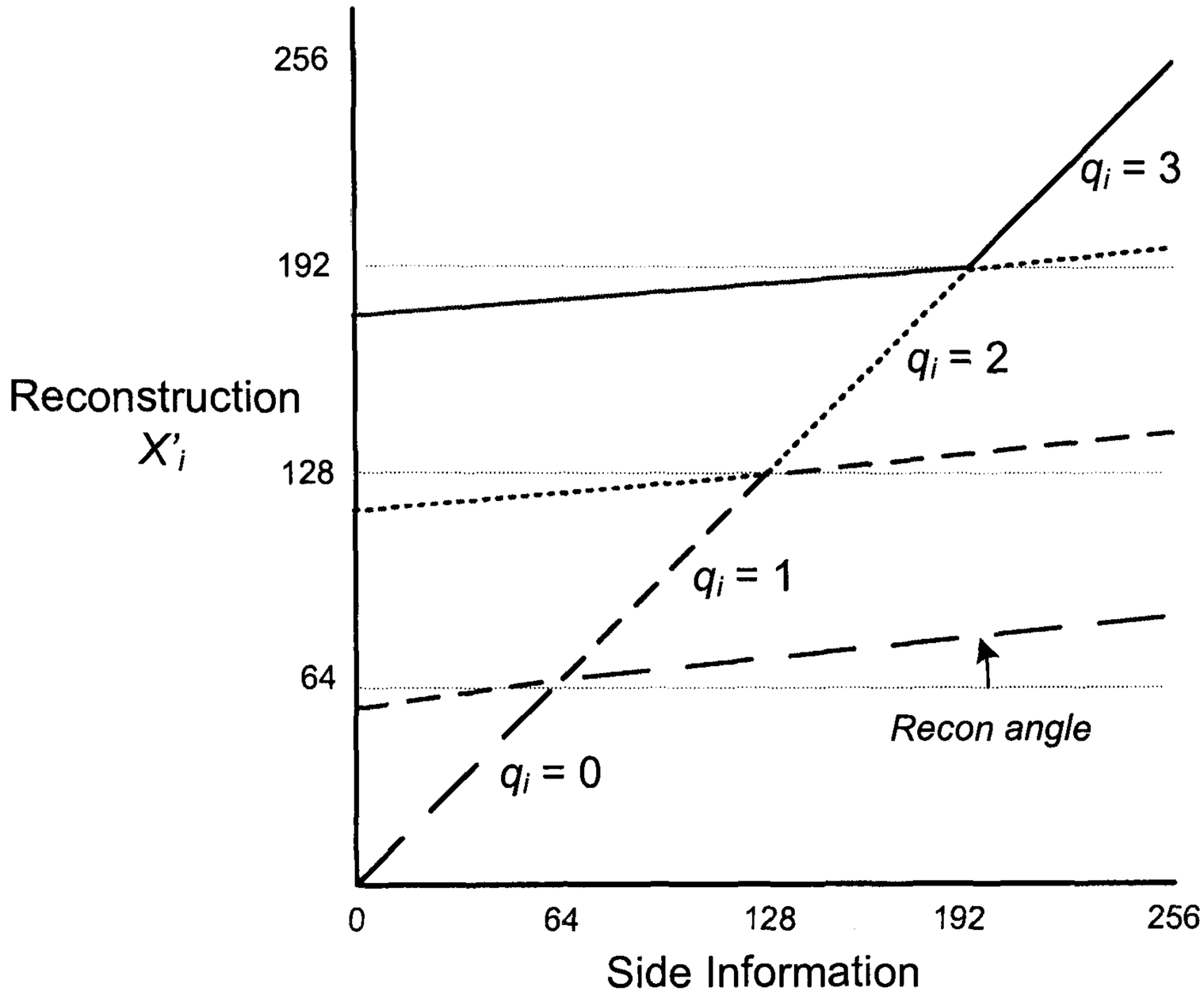


Figure 5-8: Reconstruction model for 4 quantisation levels

Optimised *recon angle* has been obtained after a training process where the PSNR is optimised by changing the *recon angle* between 0 and 45 degrees. The *recon angle* vs. BER data for each bit plane is obtained and the fitting curve is considered as the best *recon angle* (Figure 5-9) [57].

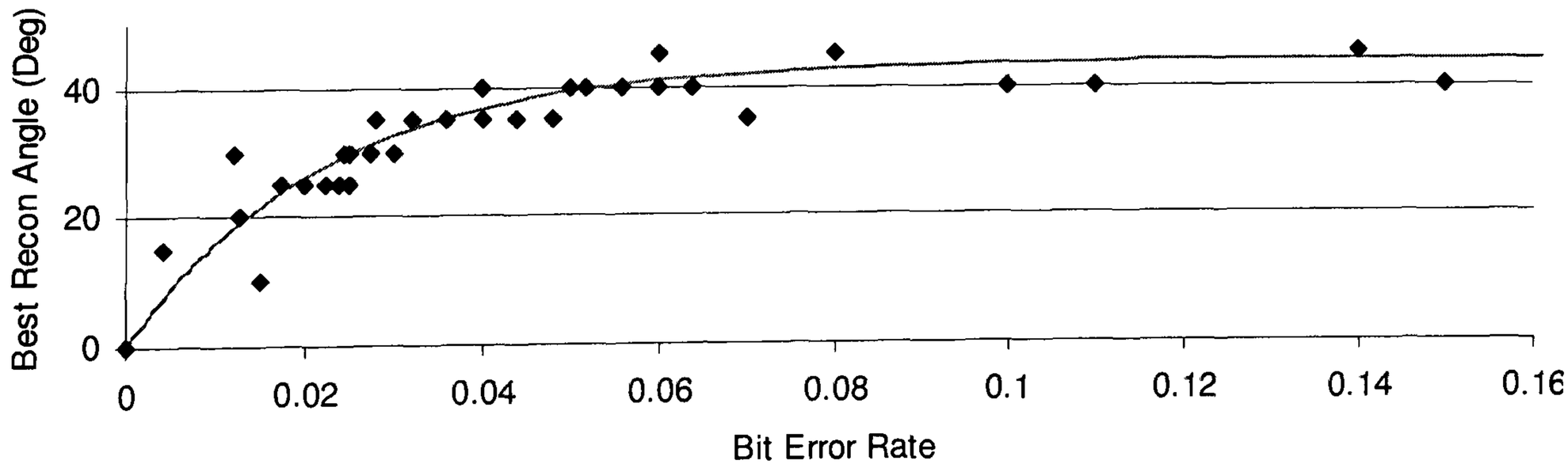


Figure 5-9: Scatter plot for the optimum recon angle and the corresponding BER (*Foreman* sequence 20 frames)

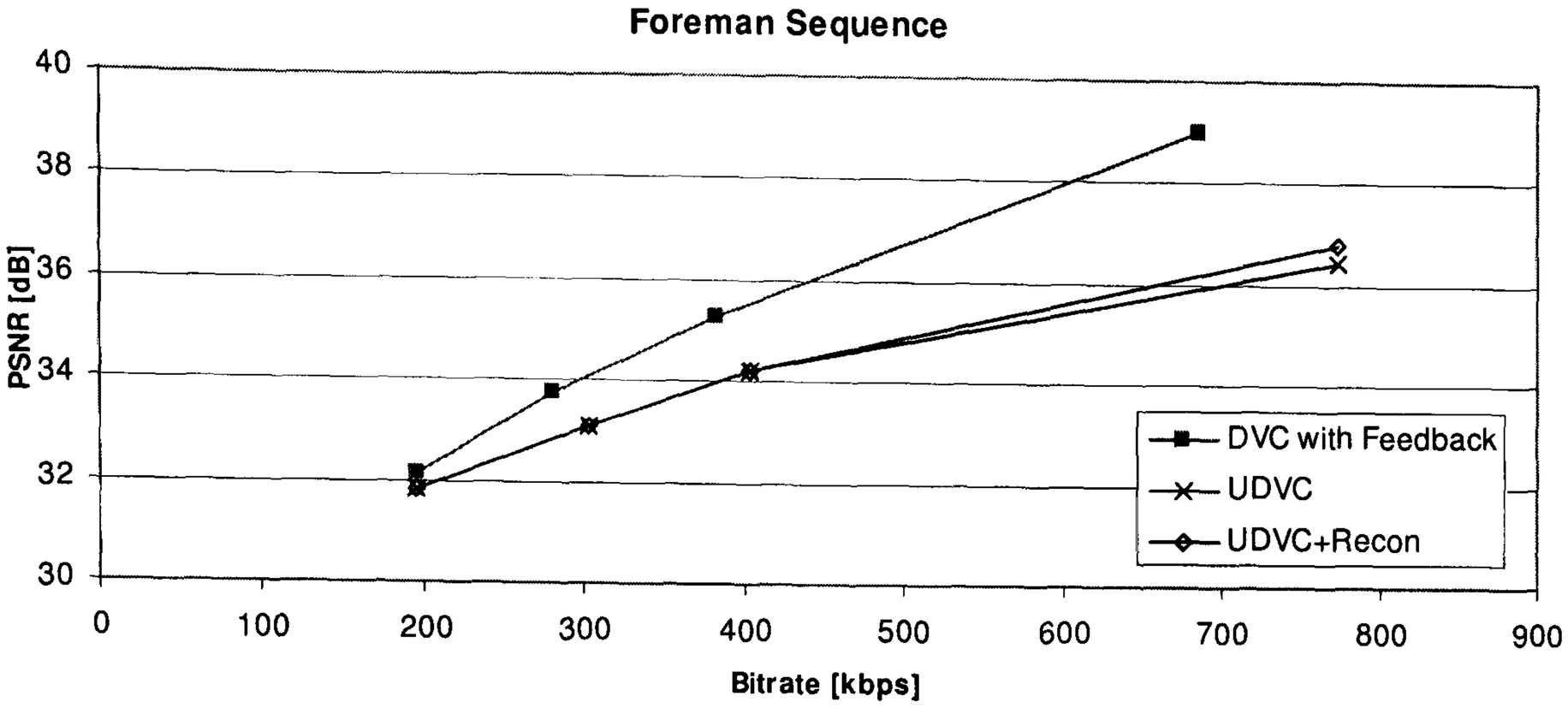
5.3.2 Simulation Results

Same test conditions are used as in the Section 5.2.3. Briefly, the proposed codec is tested for *Foreman*, *Hall-Monitor* and *Salesman* QCIF (176 × 144) test sequences. First 101 frames from each sequence are used for the simulation and only the luminance component is considered.

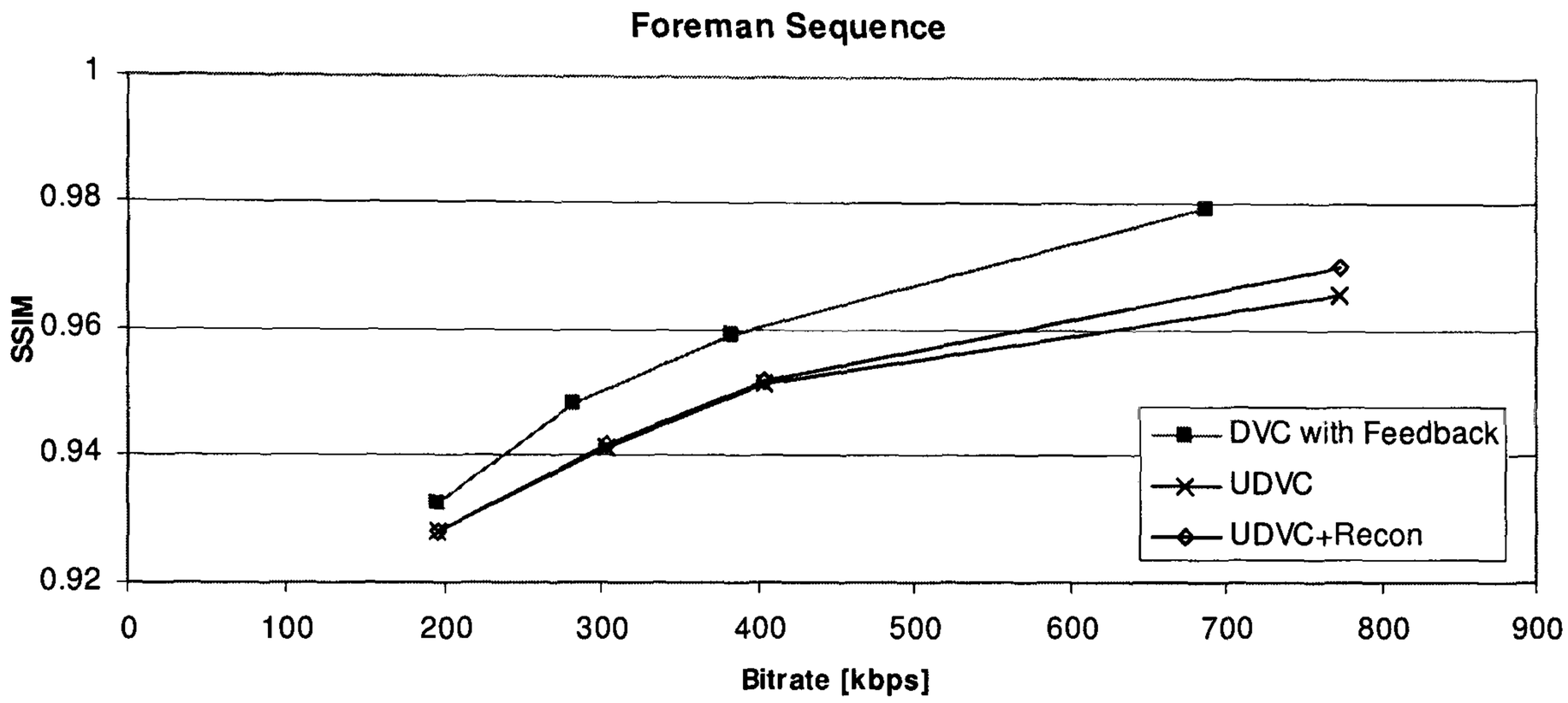
- Frame rate: 30 fps
- GOP length: 2
- DCT size is 4 × 4

SSIM and VQM quality metrics have been used as well as PSNR in order to demonstrate the subjective quality. The SSIM metric takes values between 0 and 1; the higher the SSIM index the better the quality. On the other hand, the VQM metric is mapped between 0 and 20 where the lower the VQM index the better the quality. Also examples of decoded frames are given at the end.

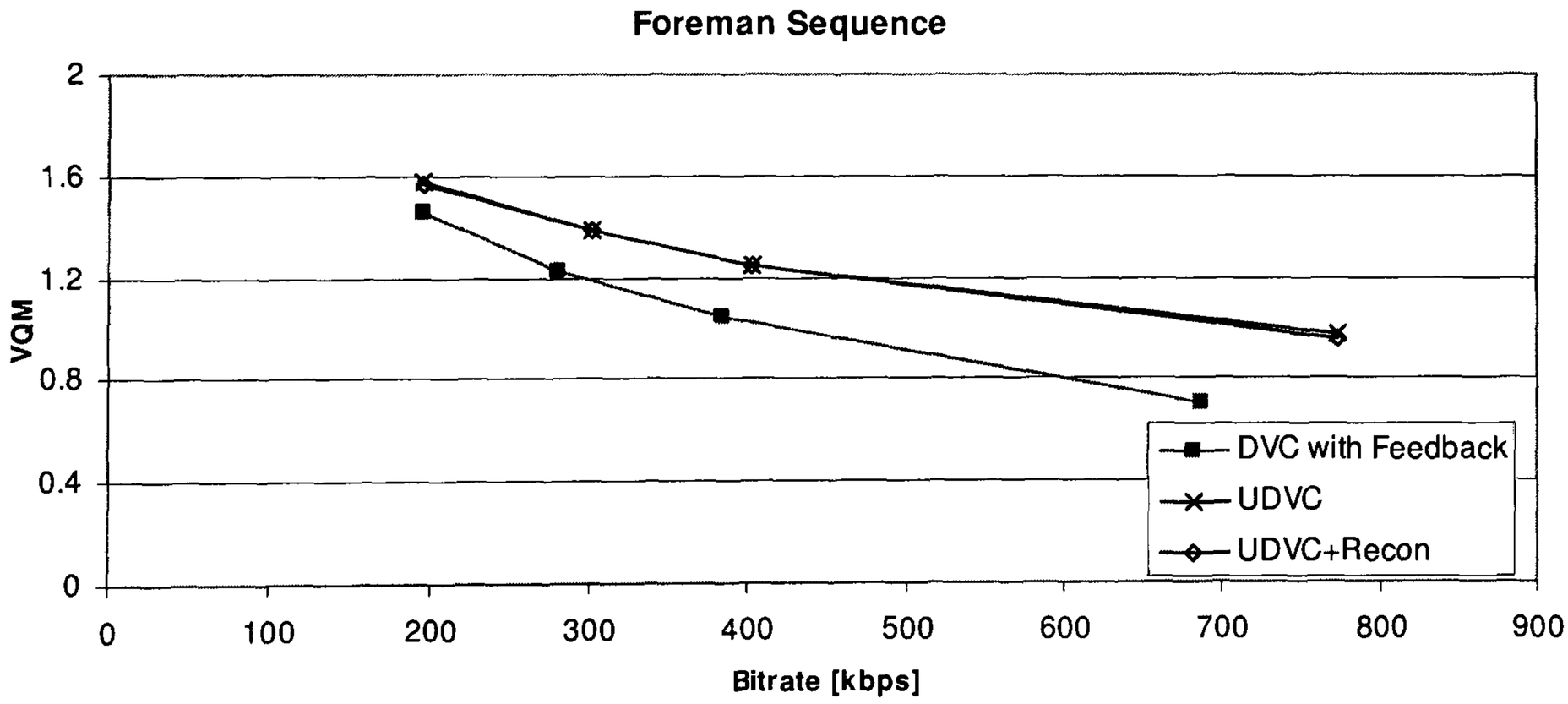
Figure 5-10, Figure 5-11 and Figure 5-12 illustrate the performance comparison of the proposed codec for the *Foreman*, *Hall-Monitor* and *Salesman* sequences respectively. The most significant improvement has been achieved for the *Foreman* sequence (Figure 5-10). When different metrics are considered, SSIM shows that the performance of the proposed UDVC codec with the improved reconstruction algorithm (UDVC+Recon) is closer to the DVC comparing against the other metrics. According to Figure 5-11 and Figure 5-12 (*Hall-Monitor* and *Salesman),* the improvement by the improved reconstruction algorithm is at minor levels when UDVC and UDVC+Recon are compared. The main reason of this is more efficient parity allocation of the UDVC codec for the *Hall-Monitor* and *Salesman* sequences when compared to the *Foreman* sequence. Therefore the BER is smaller for the *Hall-Monitor* and *Salesman* sequences, and the improvement by the improved reconstruction technique is not significant.



(a) PSNR

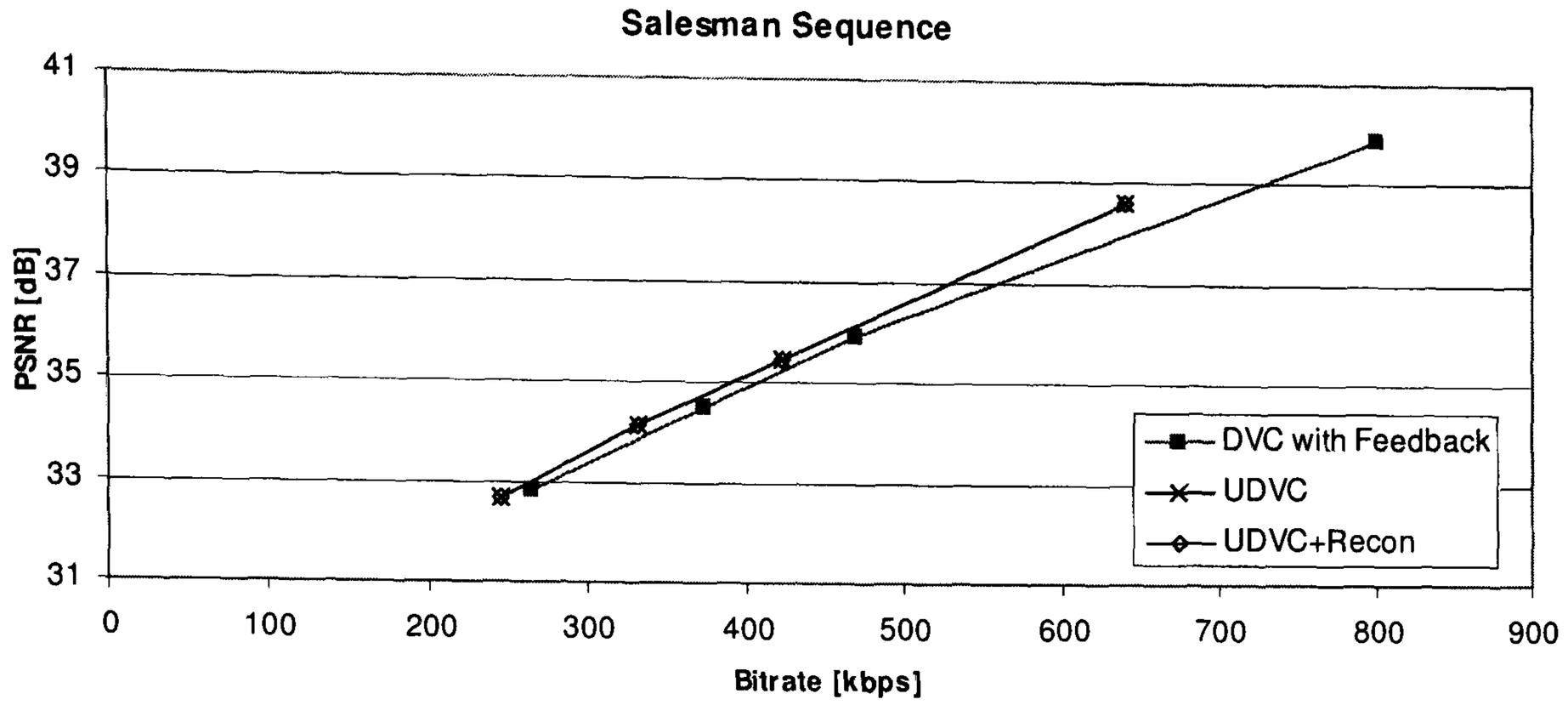


(b) SSIM

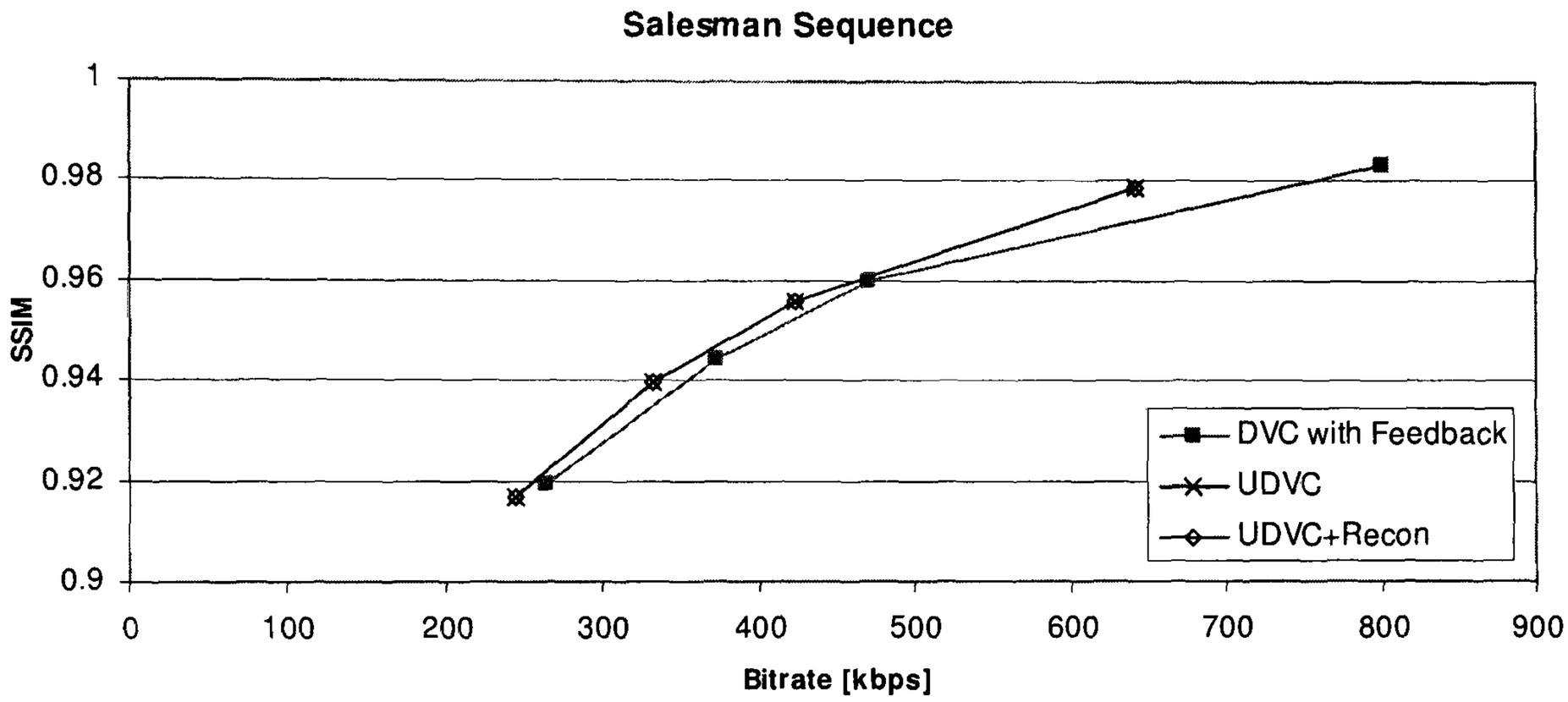


(c) VQM

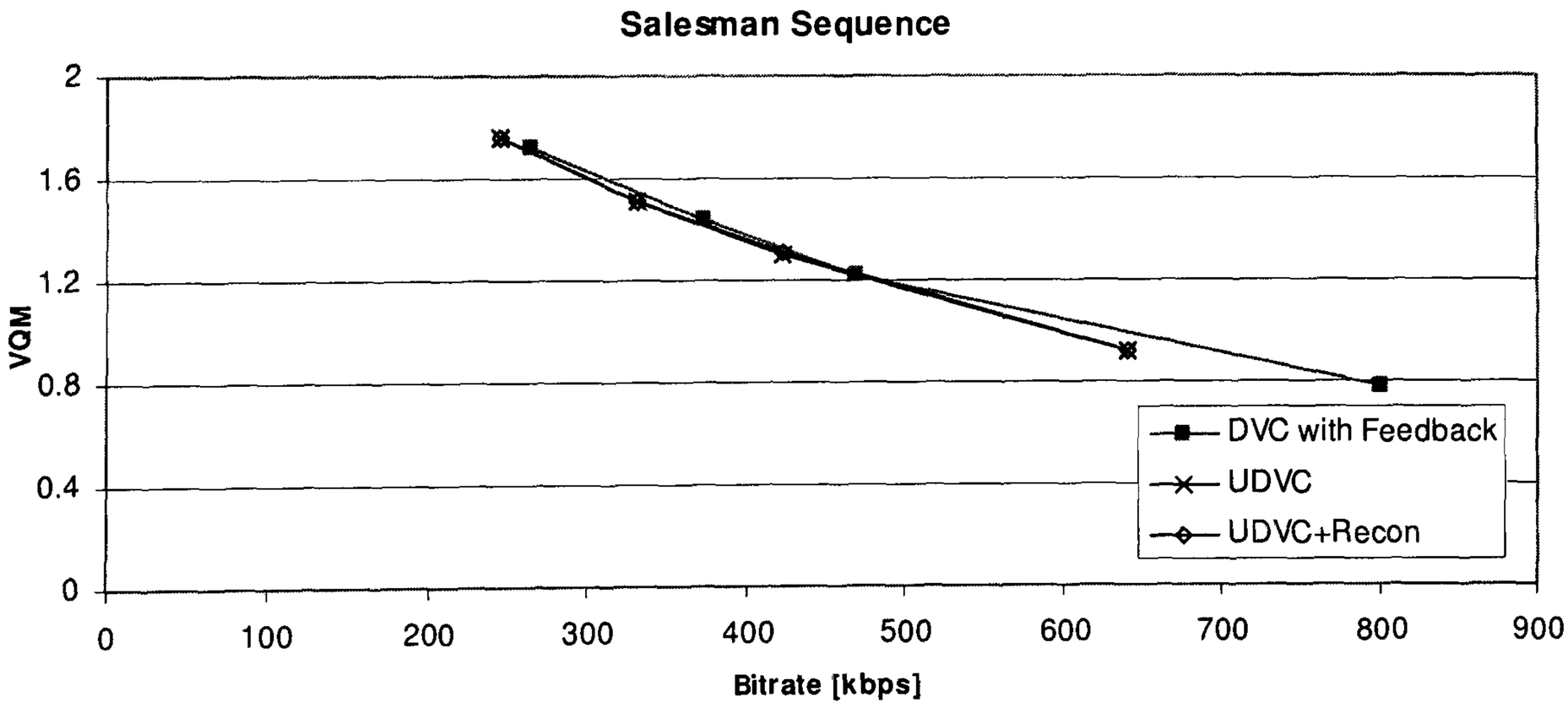
Figure 5-10: Performance comparison for *Foreman* sequence using different metrics



(a) PSNR

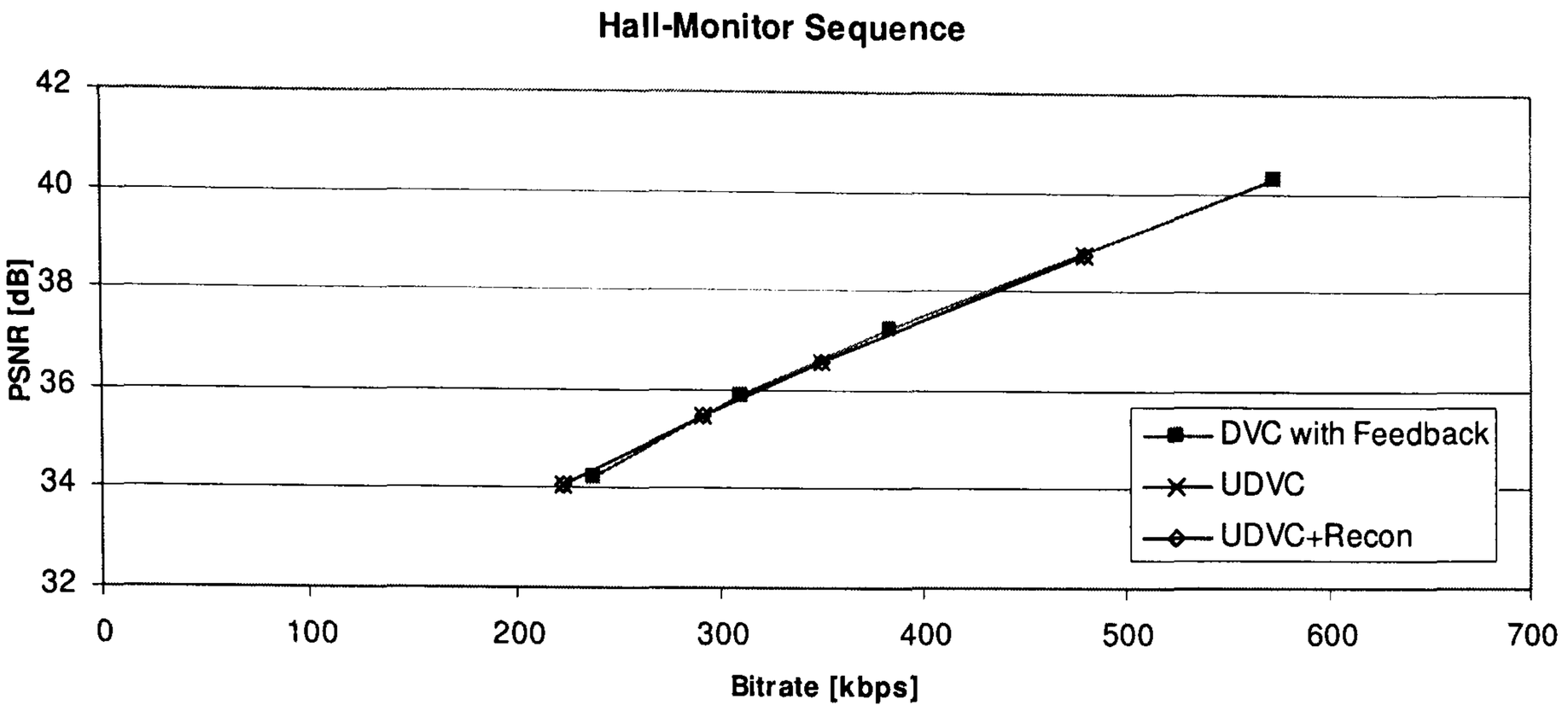


(b) SSIM

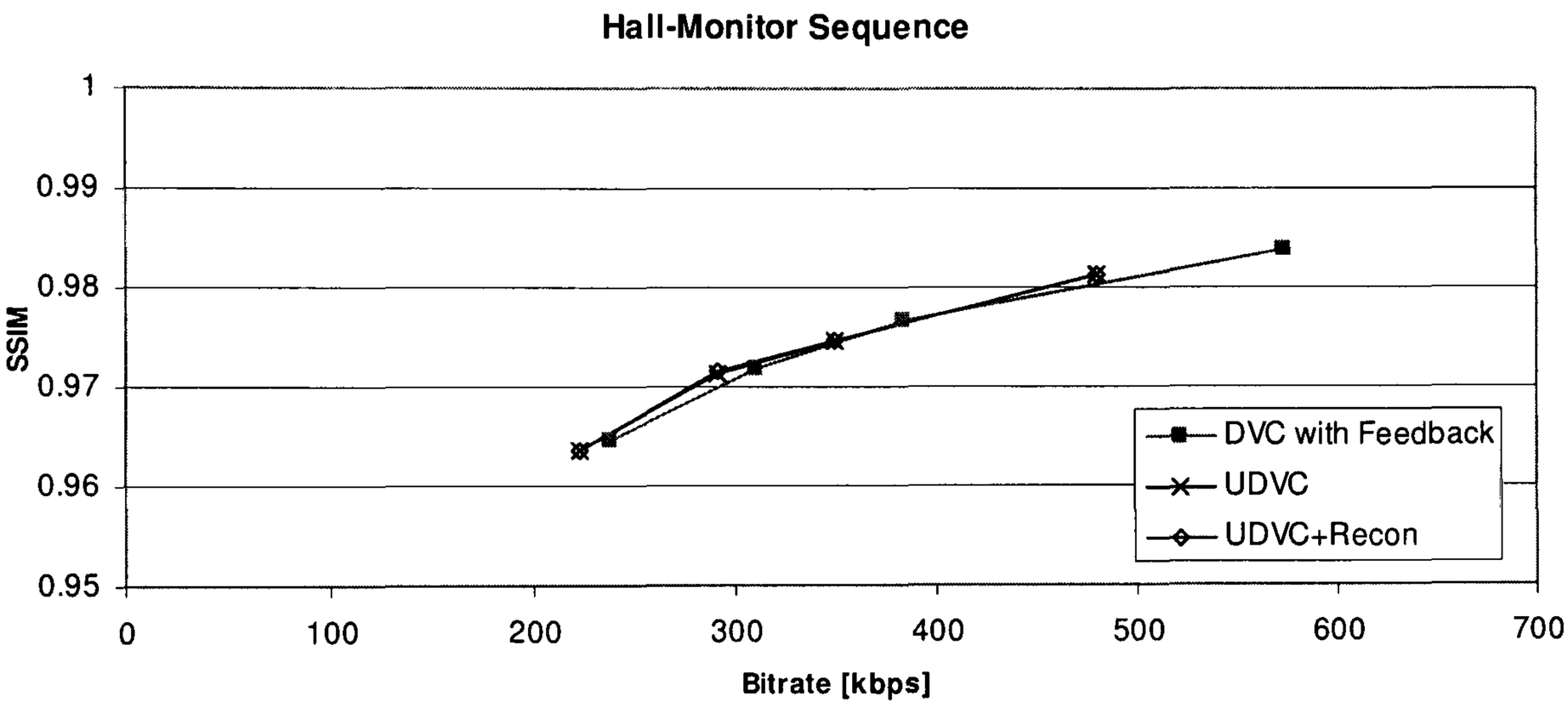


(c) VQM

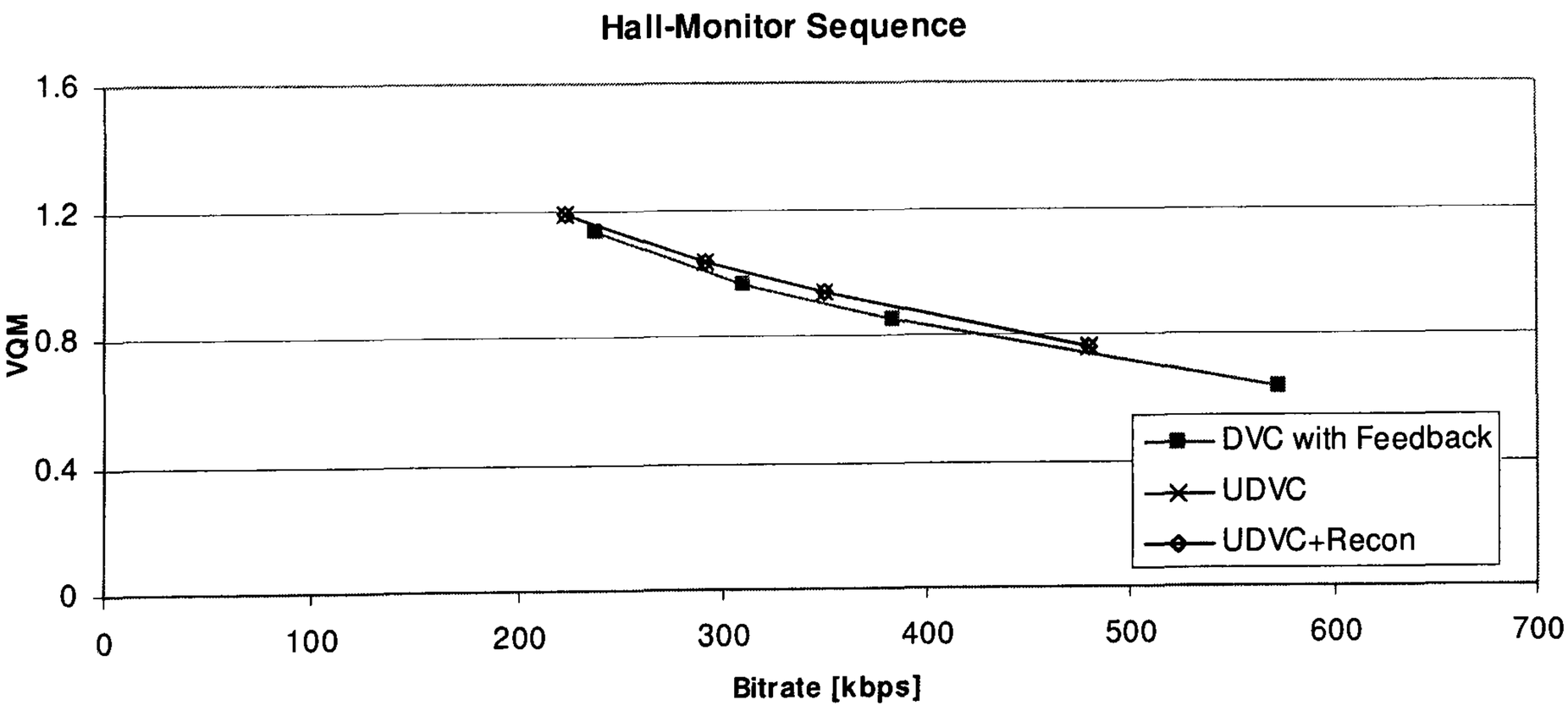
Figure 5-11: Performance comparison for *Salesman* sequence using different metrics



(a) PSNR

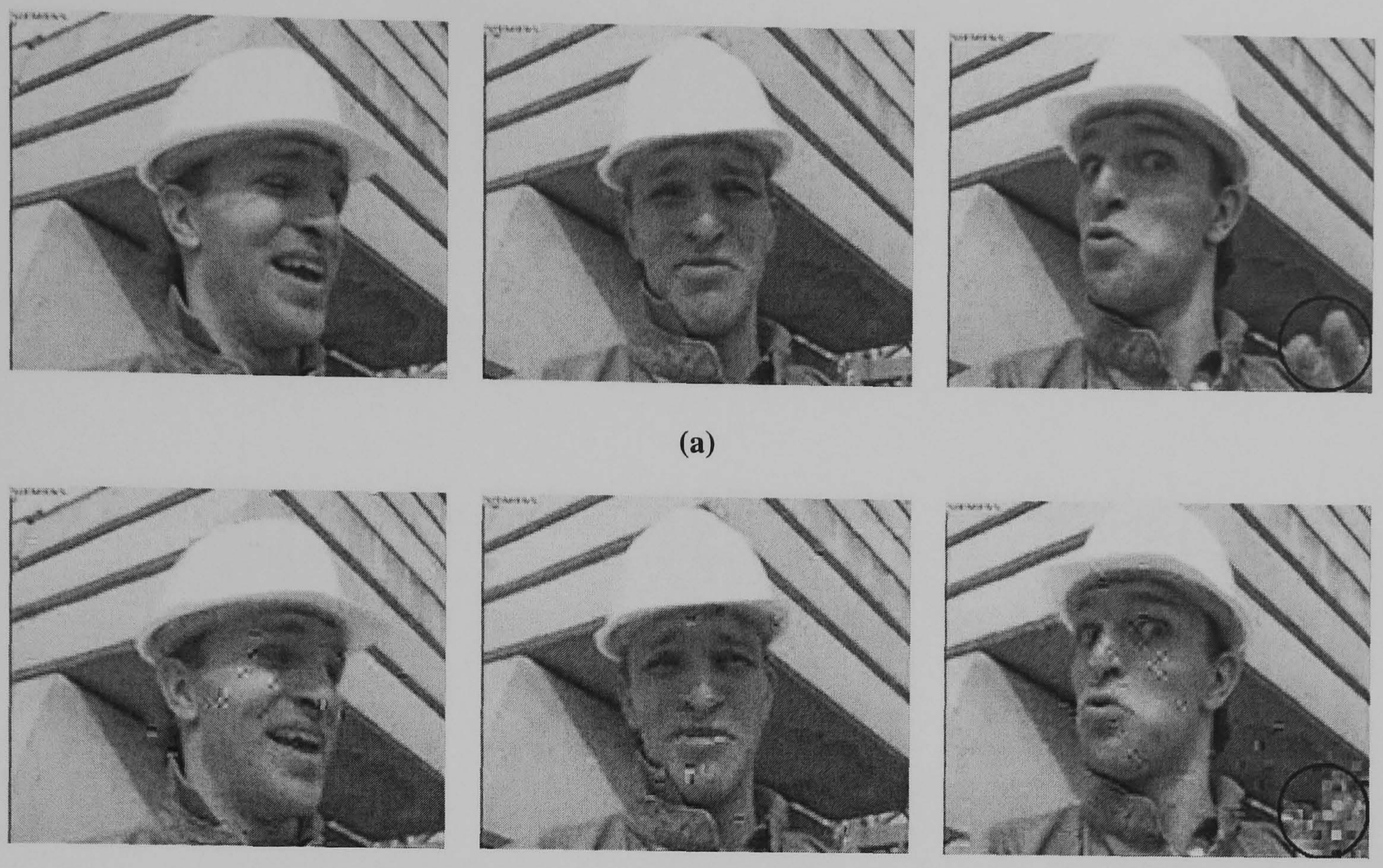


(b) SSIM



(c) VQM

Figure 5-12: Performance comparison for *Hall-Monitor* sequence using different metrics



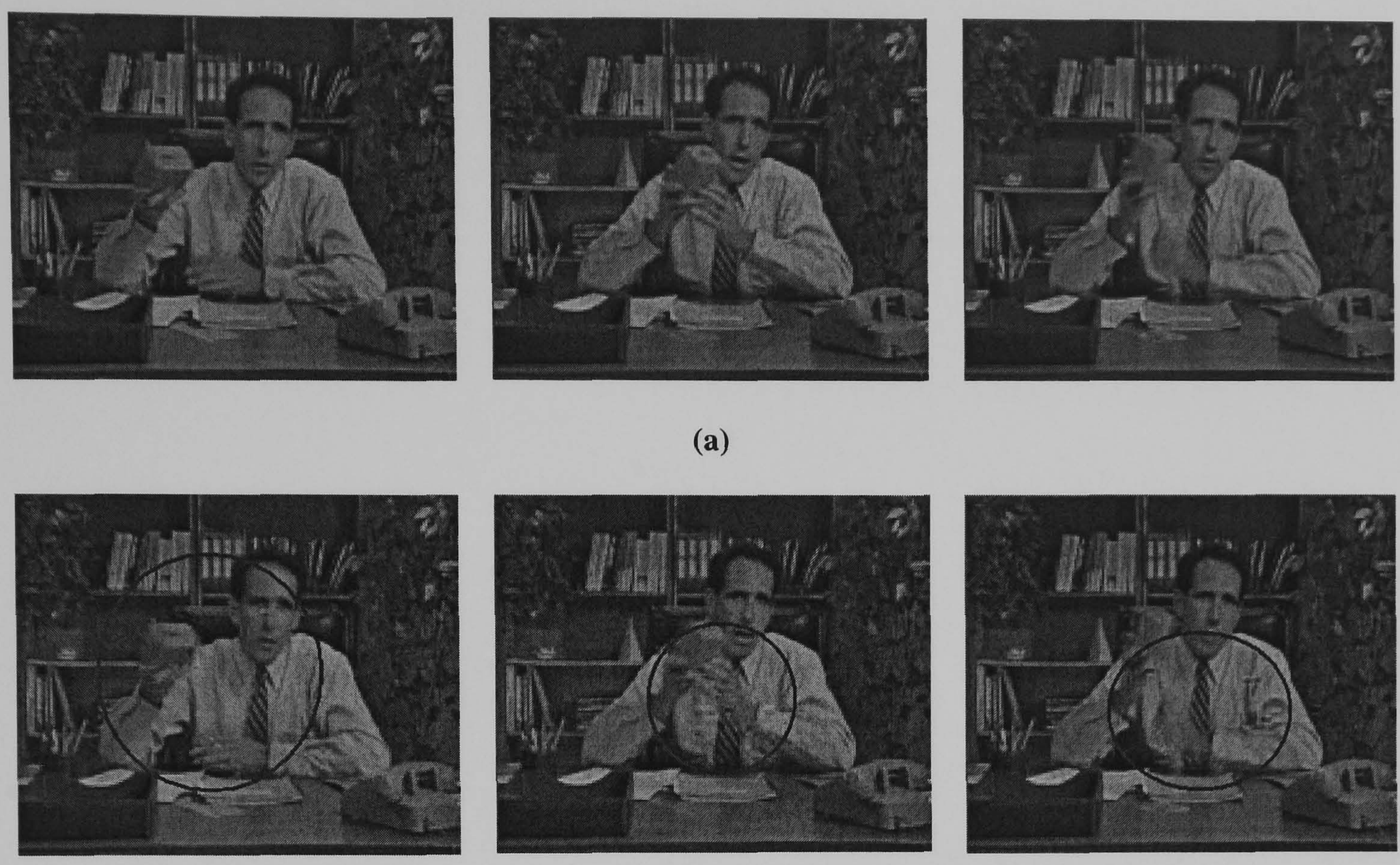
(a)

(b)

(c)

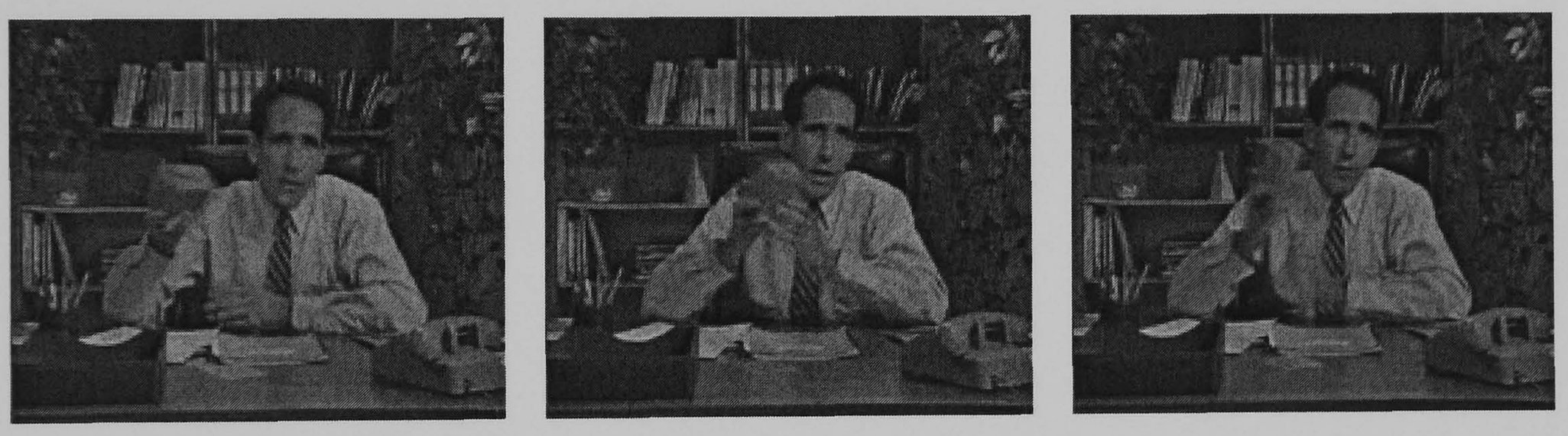
Figure 5-13: Images for the *Foreman* test sequence: (a) DVC with feedback and (b) UDVC (c) UDVC+Recon (frame number 12, 50 and 90 from left to right).

Figure 5-13, Figure 5-14 and Figure 5-15 show several frames of *Foreman*, *Salesman* and *Hall-Monitor* sequences respectively for the DVC with feedback, UDVC and UDVC+Recon codecs. Results are obtained with conditions: QM = 8 for WZ frames and QP = 26 for key frames (same as in the previous section). It can be easily seen that improved reconstruction algorithm removes most of the blocking artefacts however some parts of the frames are not completely recovered such as the fingers in 90th frame of *Foreman* sequence (as highlighted in Figure 5-13). This can happen when both the errors in the side information and the turbo coding coincide at the same part of the frame since the proposed technique relies more on the side information when the BER increases.



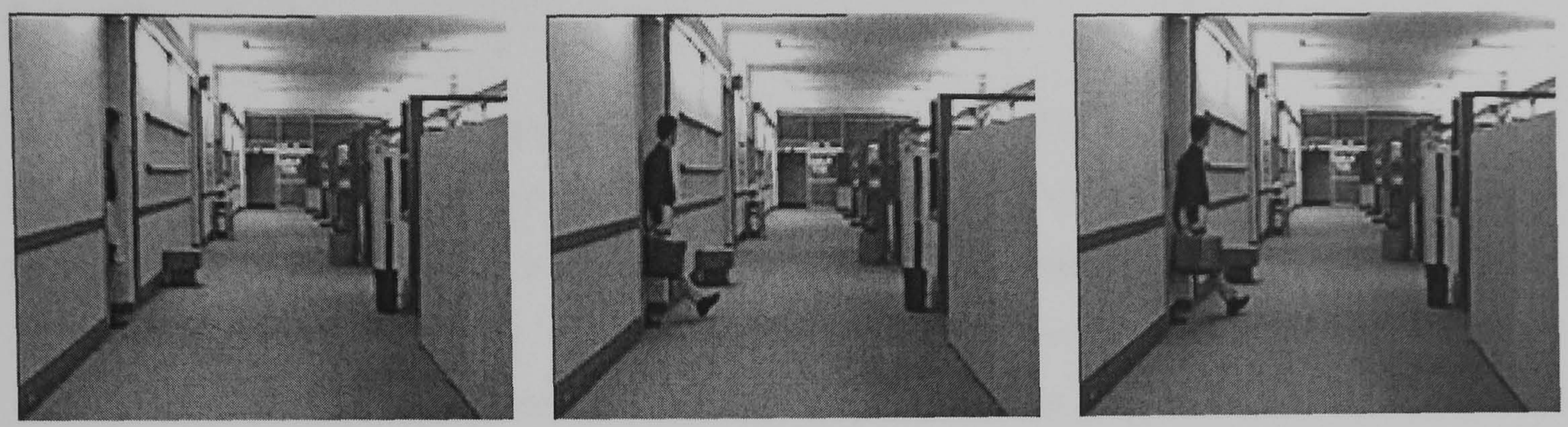
(a)

(b)

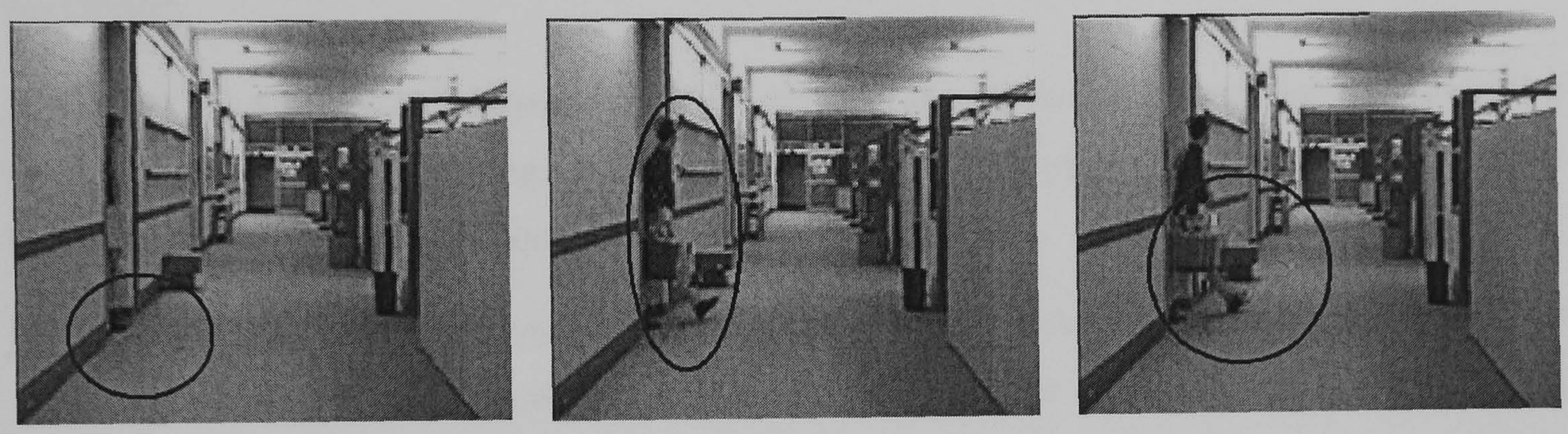


(c)

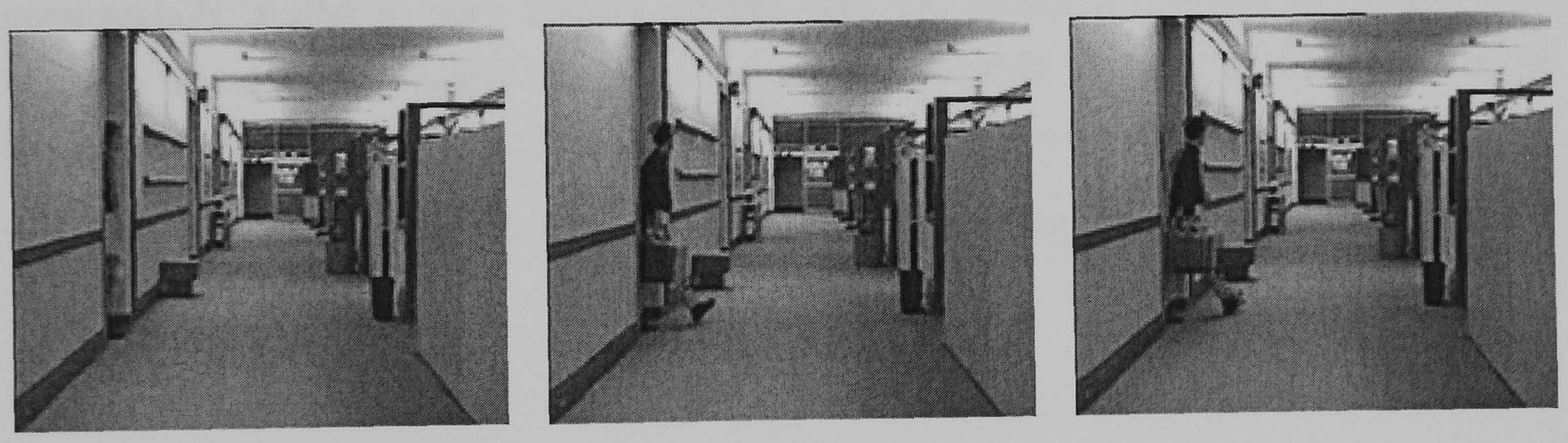
Figure 5-14: Images for the *Foreman* test sequence: (a) DVC with feedback and (b) UDVC (c) UDVC+Recon (frame number 12, 50 and 90 from left to right).



(a)



(b)



(c)

Figure 5-15: Images for the *Hall-Monitor* test sequence: (a) DVC with feedback and (b) UDVC (c) UDVC+Recon (frame number 8, 56 and 60 from left to right).

5.4 Conclusion

In this chapter, a transform domain UDVC codec is proposed. The proposed codec is a practical solution to the prevailing problem of complete dependency of DVC on a feedback channel, which restricts DVC from been used in a number of consumer electronics applications. At the encoder, parity is allocated dynamically based on the motion activity. At the decoder, these parities are used to decode bit planes. Simulation results clearly show that the RD performance proposed codec can approach that of the DVC codec with feedback, which is assumed to be the upper limit for UDVC.

In Section 5.3, an improved reconstruction algorithm is adopted in order to increase the subjective quality i.e. reducing the blocking artefacts. The results clearly show the improvement by using the improved reconstruction technique.

Potential applications of the proposed UDVC codec would include low cost video cameras where the compressed video payload is written into storage media for offline processing. The cost of cameras will be lowered by replacing the high-complexity video encoders in the existing commercially available video camera designs with the proposed dramatically simple DVC encoder. Another particularly attractive potential application domain is identified as the Disposable Video Camera, in which case production cost of the encoder is a major concern.

The proposed unidirectional DVC technique in this chapter have been published in a journal (please see appendix B).

Chapter 6

6 Modifying Frame Coding Structure

As discussed in the second chapter, DVC has attracted the interest from many researchers. Several solutions have been proposed at modular level including the ones presented in this thesis. However, the rate distortion performance of DVC is still behind that of conventional video coding (namely H.264/AVC), which is the main motivation behind the developments presented in this chapter.

In this chapter two different frame coding structures have been proposed for transform domain DVC in order to improve the rate distortion performance. First one is interlaced coding, which makes use of the strong spatial correlations between the parts within each frame. This technique reduces the gap between the conventional video coding and DVC where the temporal correlations are low, i.e. for high motion sequences.

In the second part, a block based coding architecture is given. The proposed algorithm estimates the motion activity for each block and re-orders the blocks according to the estimated motion level. Therefore, the blocks with equivalent motion levels are coded together while saving many bits for the stationary (low motion) blocks.

The rest of this chapter is arranged in the following order: In section 5.1, proposed interlaced DVC codec is presented; in section 5.2, block based coding structure is explained and finally section 5.3 concludes this chapter.

6.1 Interlaced Coding

Most DVC codec designs generate the side information utilising temporal interpolation at the decoder. For high motion sequences, the rate-distortion performance of current DVC solutions are therefore far below than that of conventional video coding techniques. The main reason for this problem is the temporal predictions performed at the DVC decoder where it does not have access to the original frame. In this section a novel sub-frame DVC technique is proposed to address the above problem. It is based on separating each frame into two sub-frames then encoding them separately and at the decoder generating the side information by applying spatial interpolation using the reference sub-frame.

6.1.1 Proposed Technique

The architecture of the proposed codec is depicted in Figure 6-1. In the proposed technique, unlike the most of the DVC solutions, all frames are coded independently. Therefore this is an intra-coding technique and no separate key frames are used. At the encoder, each frame is split into two sub-frames, which are coded separately. One of the sub-frames, key sub-frame, is conventionally intra-coded. And the other sub-frame, WZ sub-frame, is Wyner-Ziv encoded as in the conventional DVC i.e. DCT transformed, quantised, turbo encoded and parity bits are stored in a buffer. At the decoder, decoded key sub-frame is used for obtaining the prediction of the WZ sub-frame, and then the predicted sub-frame is also DCT transformed and fed into the turbo decoder and reconstruction block. Turbo decoder decodes the WZ sub-frame by requesting parity bits from the encoder until it is successfully decoded. Turbo decoded sub-frame is then reconstructed and inverse DCT transformed. Finally decoded WZ sub-frames and key sub-frames are merged to obtain the decoded frames X'_n.

6.1.1.1 Frame Splitting

In the proposed technique, input frames are divided into two sub-frames at the encoder. Odd lines are grouped together to form the key sub-frame and even lines are grouped together to form WZ sub-frame. Key sub-frame and WZ sub-frame are strongly correlated; they are encoded separately and decoded jointly (Figure 6-2).

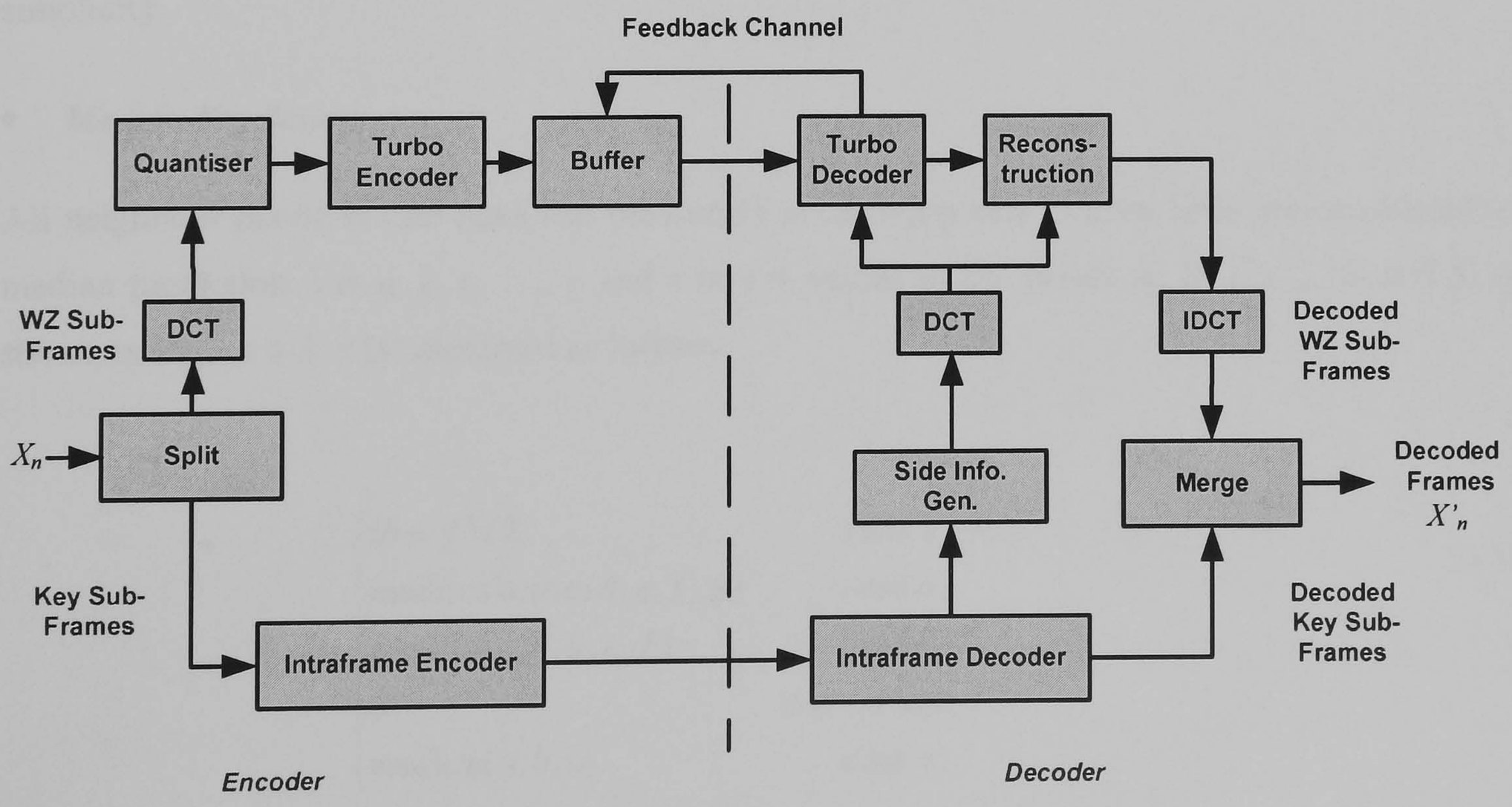


Figure 6-1: Proposed DVC codec architecture

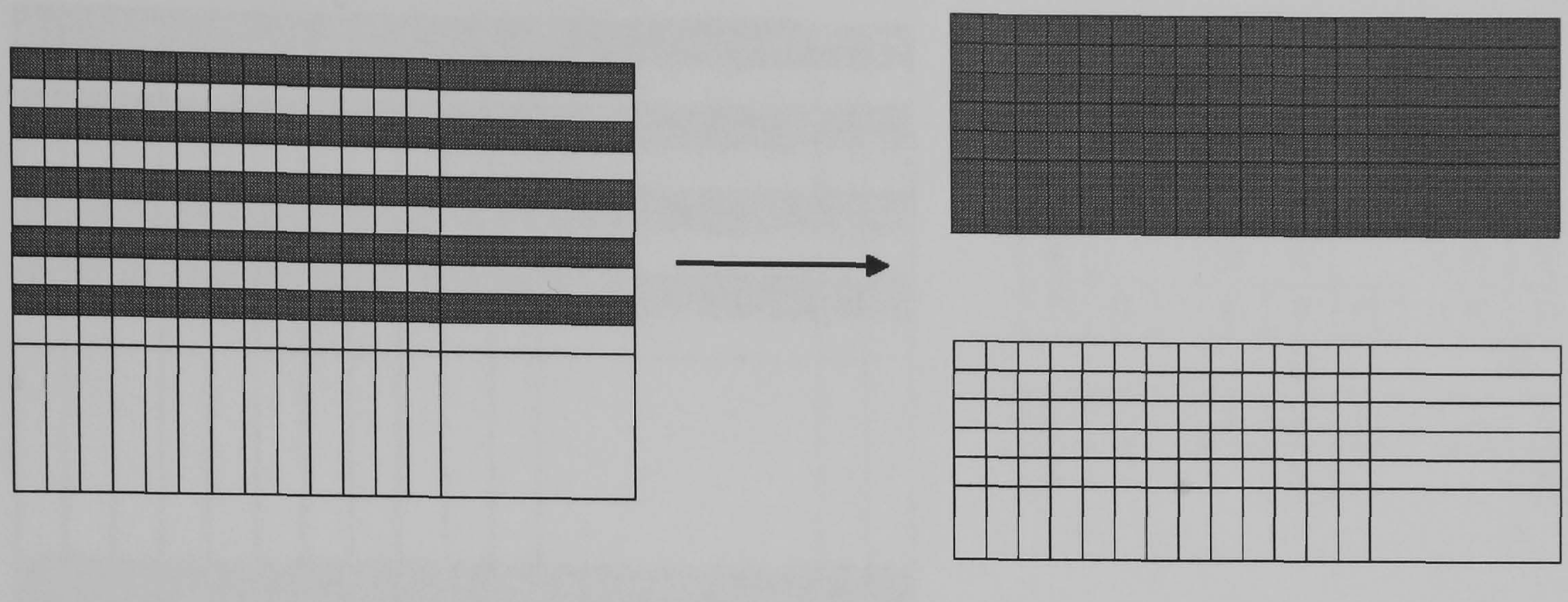


Figure 6-2: Splitting frame into two sub-frames

6.1.1.2 Side Information Generation

The proposed technique utilises only spatial prediction to obtain the side information. Main reason for this approach is the strong correlation between the two sub-frames and having a low delay coder. Once the first sub-frame is decoded, the side information is generated from the decoded sub-frame using spatial feature extraction techniques: median prediction or pixel interpolation.

As mentioned in Section 2.1.2.1 (page 14), the neighbouring pixels have higher correlations with the considered pixel. Median prediction and pixel interpolation techniques are selected in order to exploit the correlations between each pixel and its neighbours because of their computational simplicity.

- *Median Prediction:*

All neighbour pixels in odd lines and previously predicted pixels in even lines are considered in median prediction. Let a, b, c, …, g and x be the values of the pixels A, B, C, …, G and X as shown in Figure 6-3. x is calculated as follows:

$$x = \begin{cases} (b+f)/2 & case\ i \\ median(a,b,c,d,e,f,g) & case\ ii \\ median(a,b,d,e,f) & case\ iii \\ b & case\ iv\ and\ vi \\ median(a,b,c) & case\ v \end{cases} \tag{6.1}$$

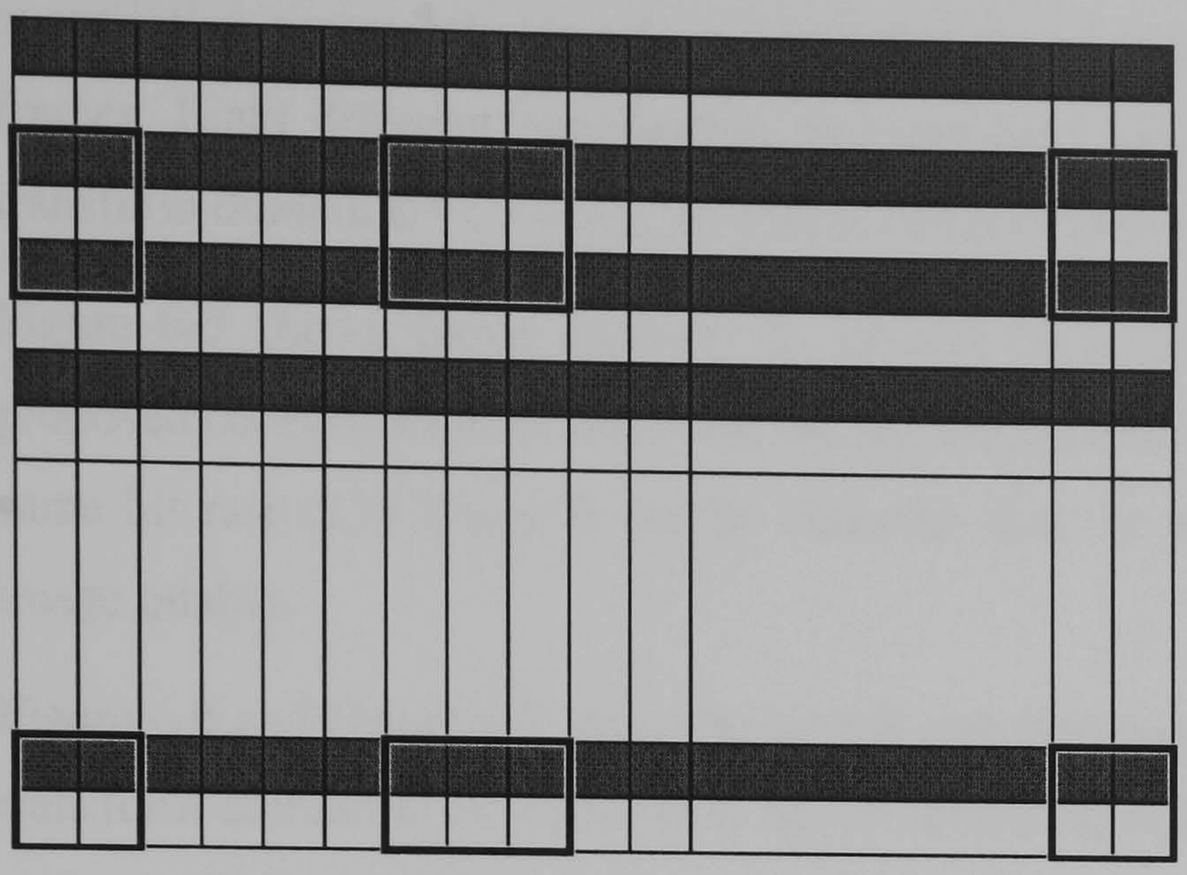


Figure 6-3: Neighbouring pixels used to predict x – median prediction

- *Pixel interpolation:*

In this technique, each pixel in the even lines is predicted as average of two neighbouring pixels in the same column, one from the previous odd line and one from the next odd line (Figure 6-4, *i*). However, the last line is directly copied from the previous odd line, since there is not an odd line after (Figure 6-4, *ii*).

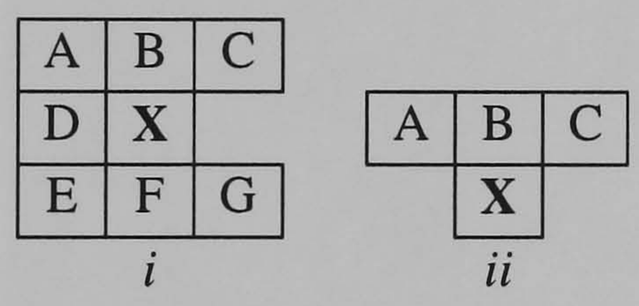


Figure 6-4: Neighbouring pixels used to predict x – pixel prediction

$$x = \begin{cases} (b+f)/2 & case\ i \\ b & case\ ii \end{cases} \tag{6.2}$$

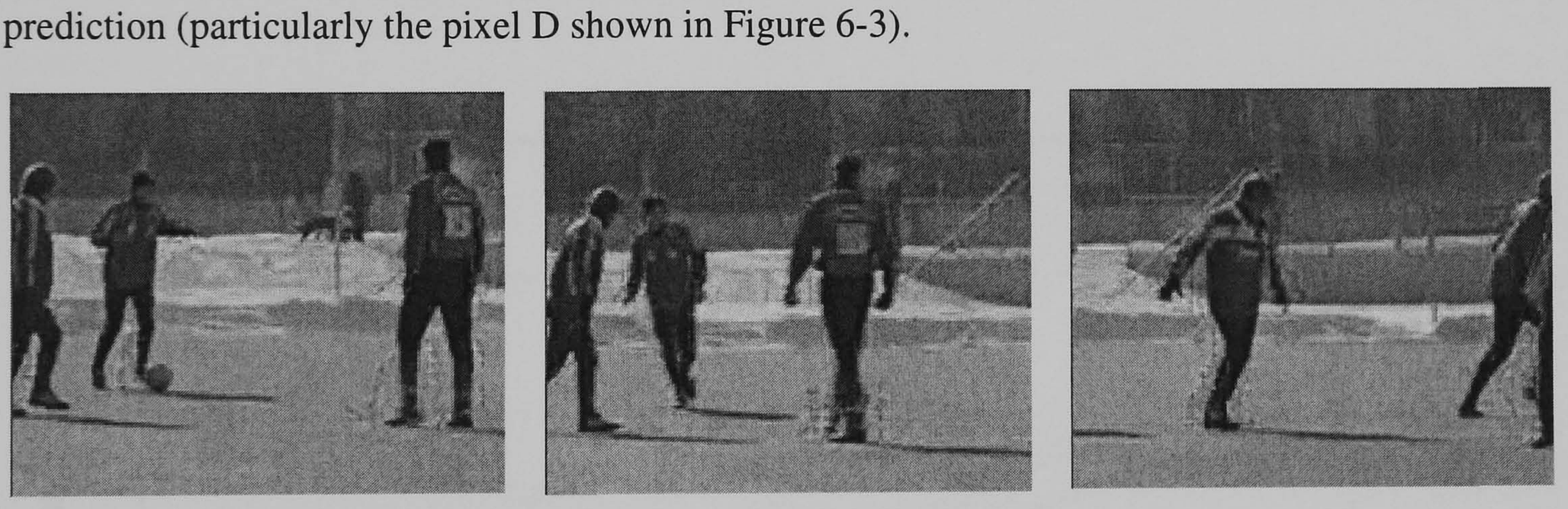
6.1.2 Experimental Results

The technique presented in this section is implemented on the codec used in [18]. Therefore, the rate distortion performance of the proposed technique has been compared against the technique presented in [18]. The performance of the proposed DVC codec is tested for 176×144 and 256×192 video sequences at 15 fps. In the experiments only the luminance component is considered and DCT size is 4×4. Key sub-frames are intra-coded using H.264/AVC and

quantisation parameters are selected to match the average PSNR of WZ sub-frames and key sub-frames. Eight different quantisation matrices have been considered, which are widely used in transform domain DVC codecs, to obtain different rate-distortion points [18].

Figure 6-5 shows frame numbers 2, 10 and 46 of *Soccer* test sequence respectively for the proposed codec with pixel interpolation and the transform domain DVC codec used in [18] at the same bit rate (220 kbps). It can be observed that the proposed codec has clearly improved the image quality.

Figure 6-6 and Figure 6-7 show the overall rate distortion performance compared to the reference transform domain DVC codec [18] and H.264/AVC with IBIB… coding structure for the *Soccer* and *Breakdancers* (view 0) sequences. It is observed that the proposed technique increases the PSNR and requires less information from the encoder. As seen in Figure 6-6, there is an average gain up to 1.5 dB for the proposed pixel interpolation technique compared to the reference [18] and the performance gap between the conventional video coding techniques and DVC is narrowed. When the proposed pixel interpolation and median prediction techniques are compared, it is seen that the former results slightly better in terms of rate distortion performance. The side information generated by pixel interpolation is more accurate because the prediction relies on the decoded pixels only, where previously predicted pixels are considered as well for median prediction (particularly the pixel D shown in Figure 6-3).



(a)

(b)

Figure 6-5: Images for the *Soccer* test sequence: (a) Reference and (b) Interlaced coding (frame number 2, 10 and 46 from left to right)

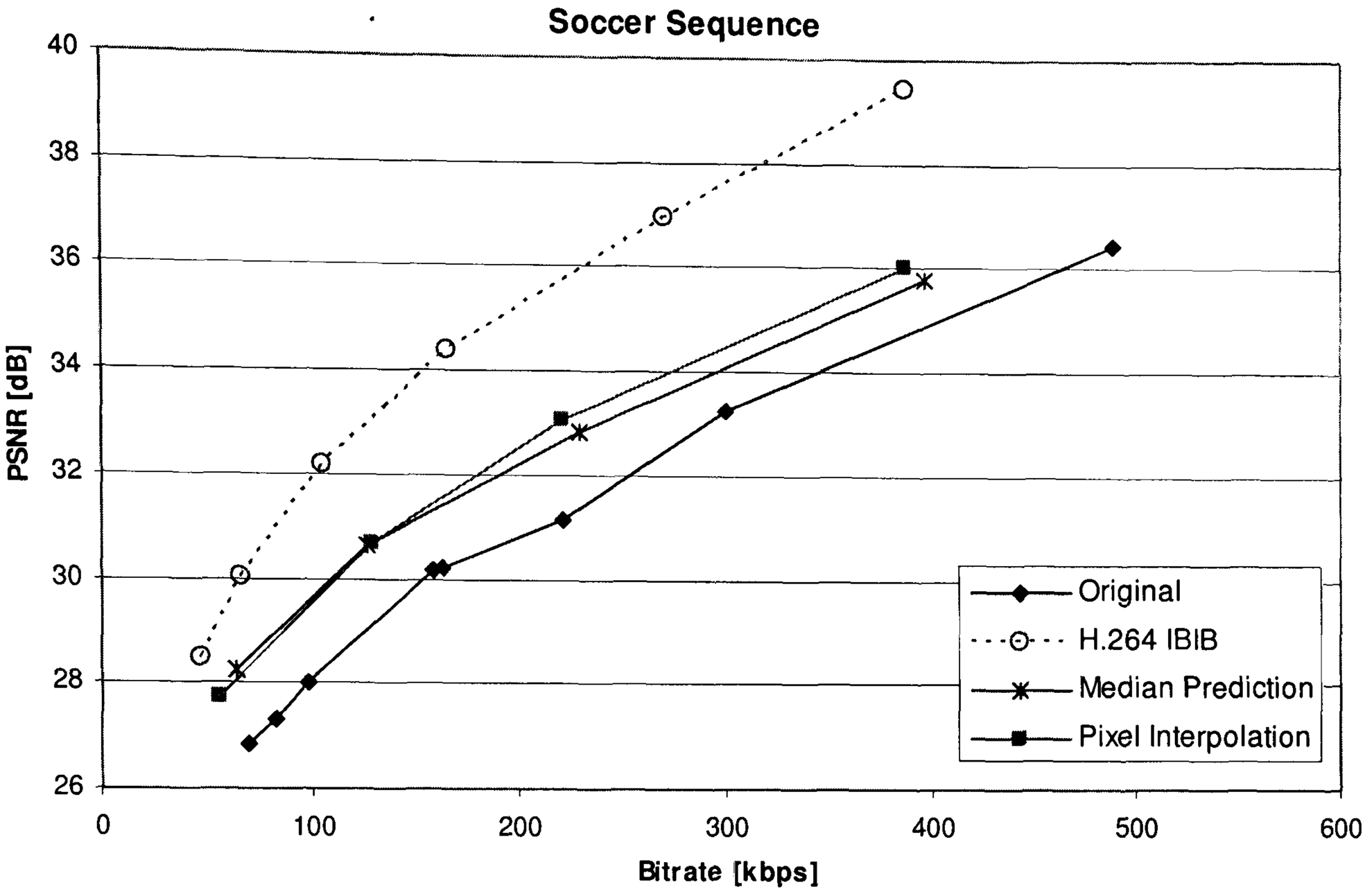


Figure 6-6: Rate-distortion performance comparison of different codecs for the *Soccer* sequence

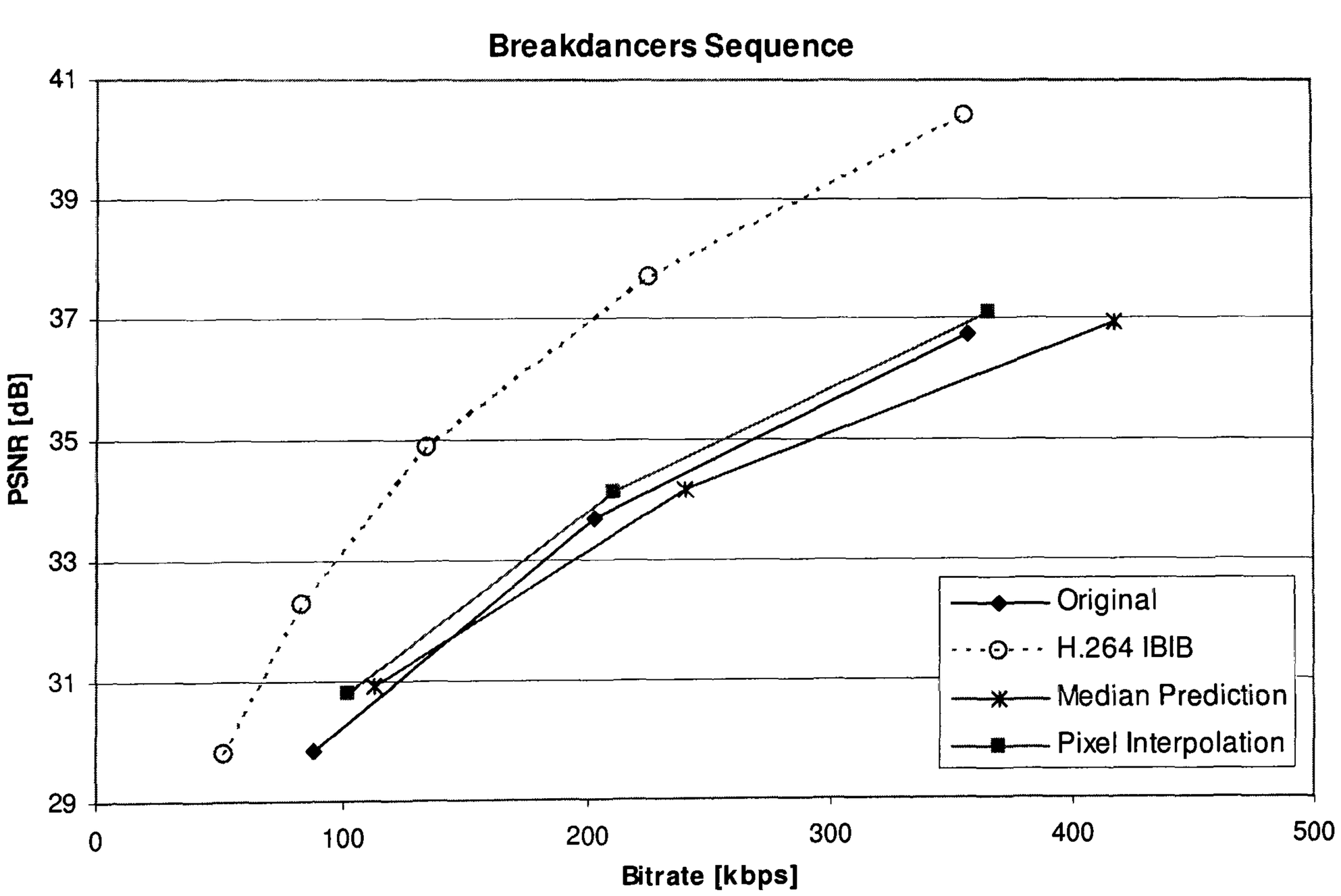


Figure 6-7: Rate-distortion performance comparison of different codecs for the *Breakdancers* sequence

6.2 Block Based Coding

In the current DVC frame work involving a feedback channel [1], the turbo encoded bit stream (parity bits) is stored in a buffer, and transmitted to the decoder upon request in small amounts. However, the puncturing process may not be efficient, particularly when the motion activity is localised in the video frame. This is a very important factor since fixed cameras with localised activity in the target scene is a very common scenario in many applications such as video surveillance systems and pc cameras. The localised motion results in the error distribution in the side information to be far from a random distribution, which is a basic consideration in the design of turbo codes. Thus the utilisation of punctured parity bits is sub-optimal. In this section, a novel region based coding scheme is proposed to exploit the above phenomenon. The frames are split into a sequence of regions based on the activity level that reflects the error distribution in the side information. The macro blocks will be re-ordered to concatenate the unevenly distributed errors and form new blocks for turbo coding to improve the efficiency of the puncturing process. The performance of the proposed technique is verified for three standard test video sequences.

6.2.1 Proposed Technique

The architecture of the proposed transform domain Wyner-Ziv codec is shown in Figure 6-8. Let's assume that all odd frames ($X_{2i\pm1}$) are key frames and even frames are Wyner-Ziv frames (X_{2i}). Wyner-Ziv frames are DCT transformed, quantised and blocks of the frames are re-ordered before turbo encoding. Parity bits are stored in the buffer and sent to the decoder upon request. Key frames are intra-coded, and at the decoder they are used in the side information generation process. Side information is also DCT transformed to obtain a similar bit stream as at the encoder it is then fed into the turbo decoder and reconstruction block. The identification of the region(s) of activity is performed at the encoder by considering the adjacent key frames as reference frames and calculating the incremental difference of the pixels in each corresponding macro block (16×16). The Sum of Absolute Difference (SAD) is used as the metric for determining the level of activity in the scene. Then these blocks are re-ordered in the order of magnitude of SAD values and each consecutive m macro blocks are grouped together for turbo coding. This algorithm is further elaborated step-by-step in the following sections.

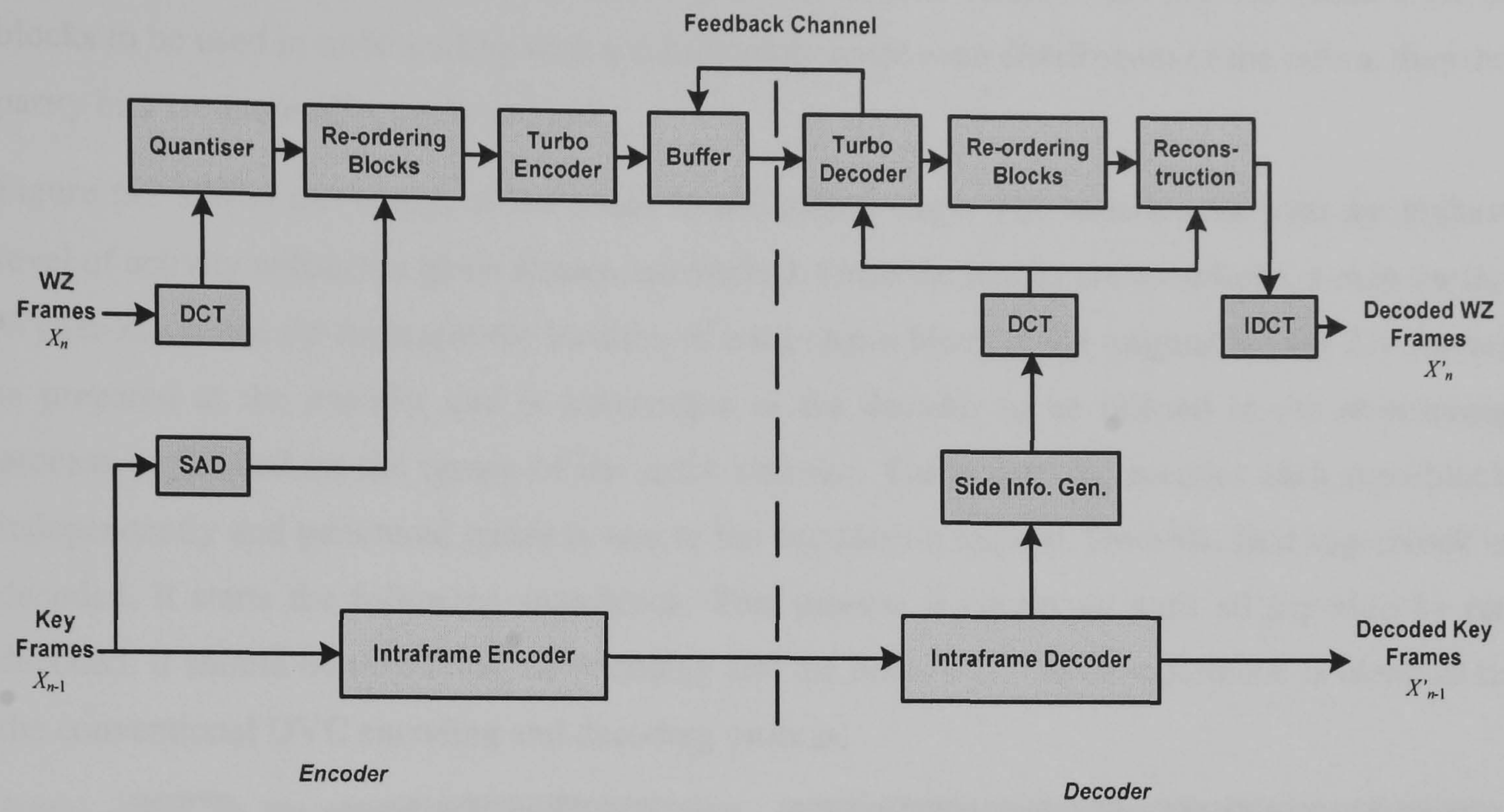


Figure 6-8: Proposed codec architecture

6.2.1.1 Locating the Region(s) of Activity

To determine the regions of activity for the Wyner-Ziv frames, the two neighbouring key frames of the Wyner-Ziv frame are considered as the references. This choice is justified by the fact that the estimation of side information at the decoder for the current Wyner-Ziv frame is utilising information from the same reference frames; hence sufficiently reflect the error distribution in the side information blocks fed to the turbo decoder. The frames are split into N regular macro blocks of 16×16 pixels. The two key frames are compared and the incremental pixel level difference between them is calculated for each macro block to estimate the motion activity. The SAD for a macro block (B) is calculated as,

$$SAD_B = \sum_B \left| X_{2i-1}(x, y) - X_{2i+1}(x, y) \right| \tag{6.3}$$

Where (x, y) represents the location of the pixels within the macro block. Summation is taken for all pixels within the B.

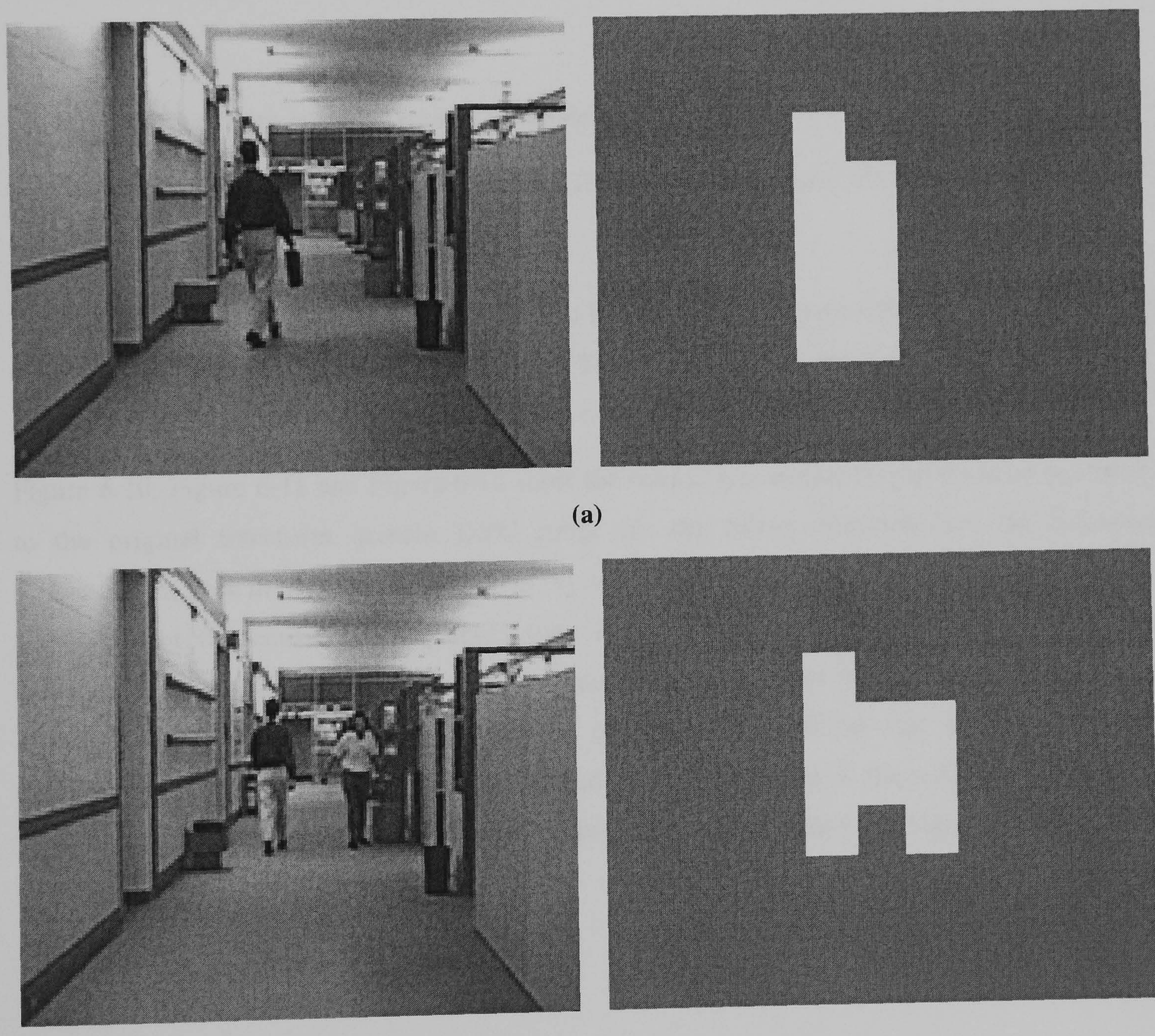
This SAD is then used as a measure of the level of activity of the macro block.

6.2.1.2 Forming the Blocks for Turbo Coding

The macro blocks are re-ordered in the descending order of magnitude of the SAD values calculated using (6.3). The macro blocks are then sub-grouped in to blocks of m consecutive macro blocks (defined as superblocks) in the re-ordered sequence. This process results in the

separation of high motion blocks from less (or zero) motion blocks. This process yields a set of blocks to be used in turbo coding with a significantly more even distribution of the errors, thus the parity bits are more effectively utilised.

Figure 6-9 shows the output of the block identification stage. The superblocks with the highest level of activity within the given frames are marked. Once the blocks are re-ordered, a map for the Wyner-Ziv frame (to represent the location of each macro block in the original Wyner-Ziv frame) is prepared at the encoder and is transmitted to the decoder to be utilised in the re-ordering process performed on the output of the turbo decoder. Turbo encoder encodes each superblock independently and punctured parity is sent to the decoder on request. Once the first superblock is decoded, it starts the following superblock. This process is continued until all superblocks are decoded. It should be noted that the encoding and the decoding of each superblock is identical to the conventional DVC encoding and decoding process.



(a)

(b)

Figure 6-9: (a) 30th (b) 90th frame of *Hall-Monitor* sequence at 15 Hz and the corresponding most active region (m = 9)

6.2.2 Experimental Results

The performance of the proposed DVC codec is tested for a number of QCIF (176 × 144) video sequences with GOP=2. The technique presented in this section is implemented into the codec used [18]. Therefore, the rate distortion performance of the proposed technique has been compared against the technique presented in [18]. The proposed codec is tested for *Silent, Hall-Monitor* and *Salesman* QCIF (176 × 144) test sequences. These sequences have been specially selected to represent a common scenario in video surveillance systems. First 101 frames from each sequence are used for the simulation and only the luminance component is considered

- Frame rate: 15 fps
- GOP length: 2
- DCT size: 4 × 4
- $m = 9$
- The key frames are considered to be available at the decoder
- Error detection mode: Ideal error detection (Bit error rate threshold: 10^{-3})

Eight different quantisation matrices (as defined in Figure 2-29 in page 42) have been selected in order to get different rate-distortion points. These quantisation matrices have been widely accepted to compare the results in the transform domain DVC codecs.

Figure 6-10, Figure 6-11 and Figure 6-12 show the overall rate distortion performance compared to the original transform domain DVC codec for the *Silent, Hall-Monitor* and *Salesman* sequences. As seen in the figures, up to 20% bit rate saving is demonstrated and this is sufficiently consistent for the considered sequences over the entire range of the rate distortion points. However, it is observed that, when very low quantisation matrices are considered, the bit rate reduction is over-shadowed by the addition of small overhead bit rate consumed by the transmission of the block map to the decoder and returns a detrimental effect to the performance. This effect is evident with the Q1 and Q2 of the *Salesman* sequence. At higher bit rates, these overhead bits become insignificant.

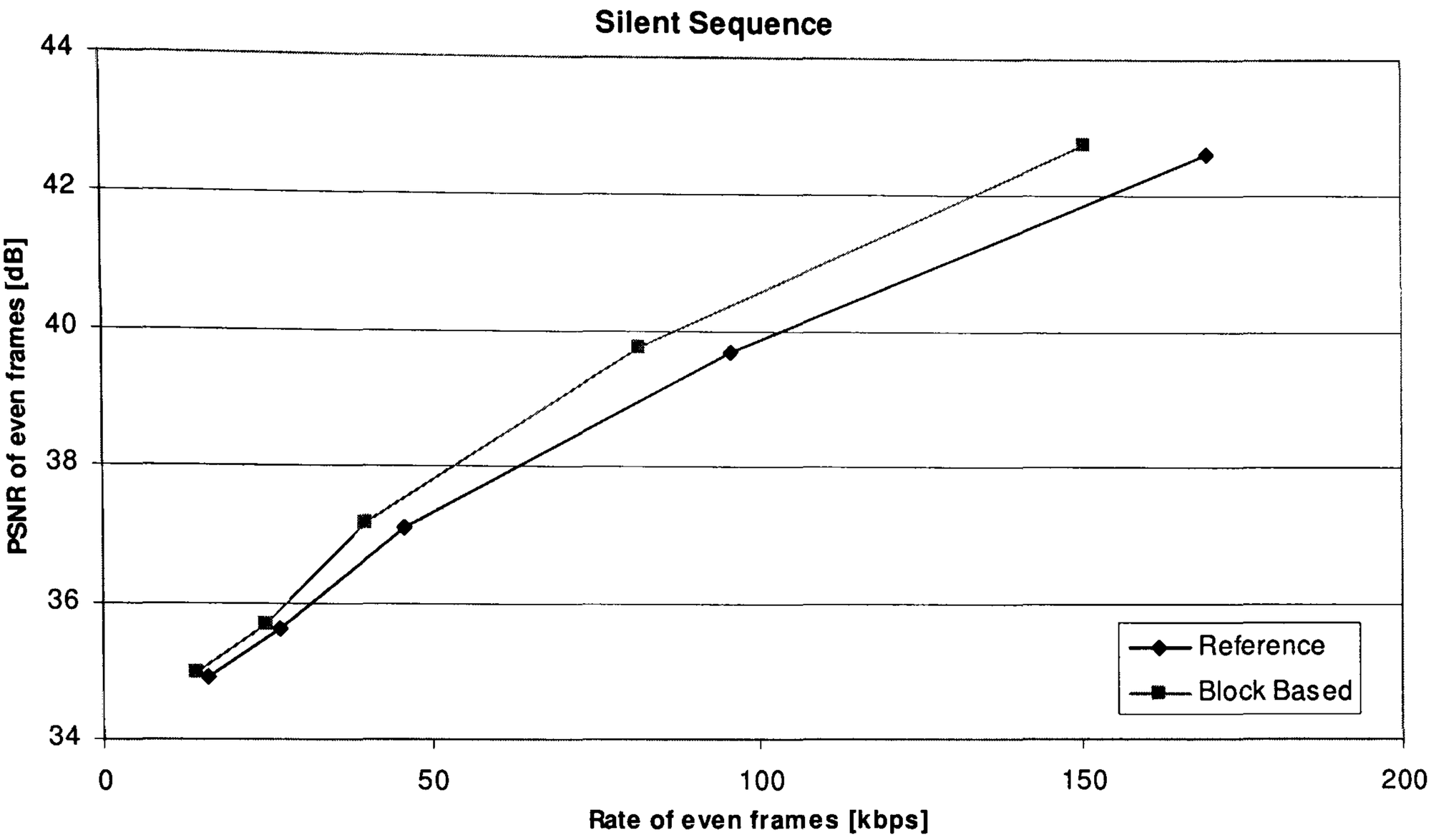


Figure 6-10: Rate-distortion performance comparison of WZ frames for *Silent* sequence

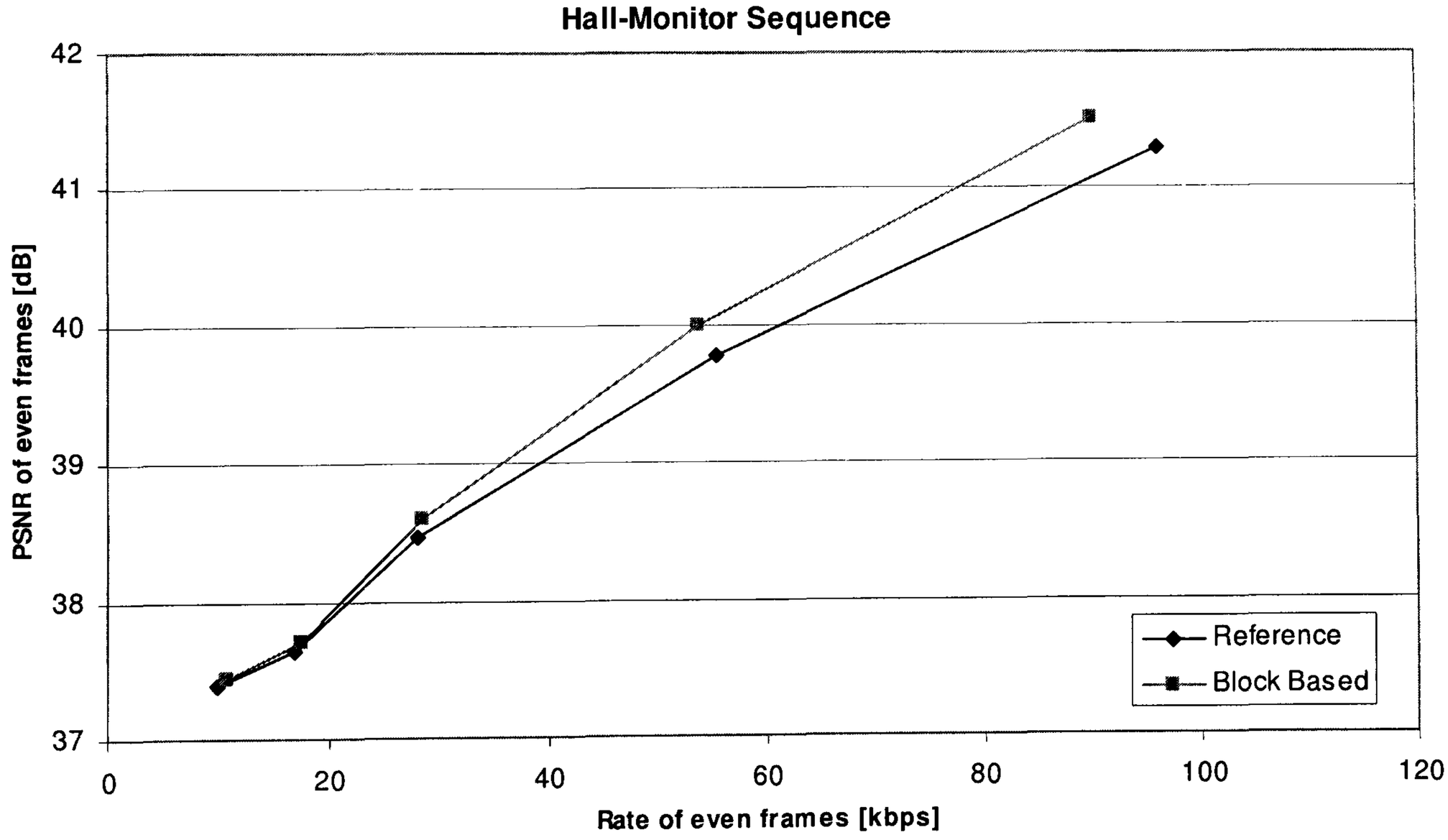


Figure 6-11: Rate-distortion performance comparison of WZ frames for *Hall-Monitor* sequence

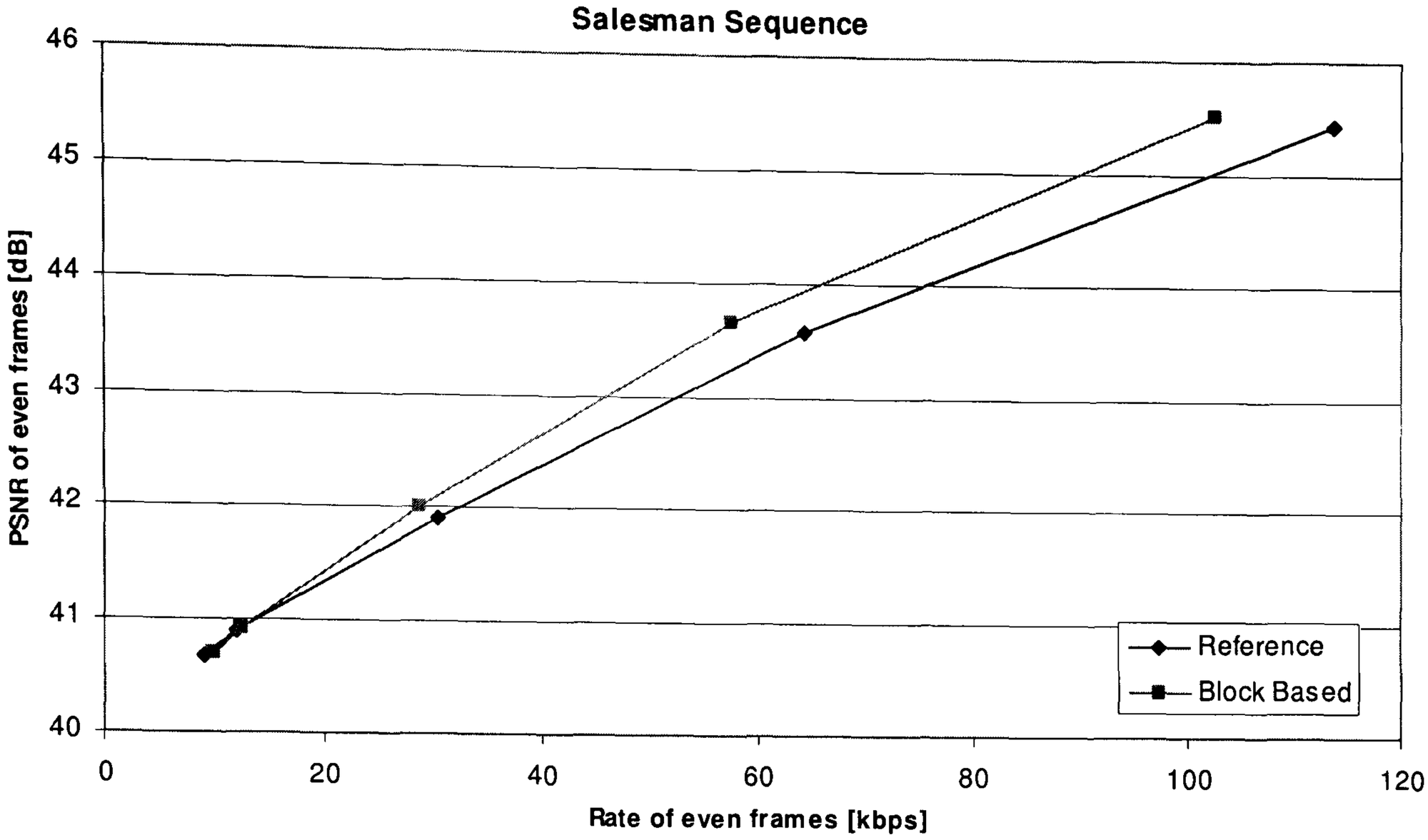


Figure 6-12: Rate-distortion performance comparison of WZ frames for *Salesman* sequence

6.3 Conclusion

In this chapter, two different frame coding structures have been proposed to improve the rate distortion performance of the transform domain DVC.

In the first section, a novel transform domain DVC codec exploiting spatial correlations at the decoder has been proposed. The proposed technique is an improved solution for especially high motion sequences. In the proposed technique, frames are divided into two sub-frames; one is intra-coded using a conventional intra-coding and the other is WZ coded. At the decoder, the prediction of the WZ sub-frame is generated by two different techniques: median prediction and pixel interpolation. Since no temporal dependency is considered in the proposed algorithm it is an intra-coding technique; therefore, the memory requirement at the encoder and decoder is significantly diminished and in addition, overall latency is reduced. Simulation results clearly show that the proposed codec outperforms the reference transform domain DVC codec [18] by a significant margin.

In the second part of this chapter, a novel technique for transform domain DVC using region-of-activity based coding. In the proposed technique, Wyner-Ziv frames are divided into macro blocks. Then, the macro blocks are re-ordered according to the level of motion activity, represented by the SAD of two neighbouring key frames. This results in a significantly better

utilisation of the parity bits yielding a performance gain in the form of a reduction in bit rate for a given picture quality. Simulation results clearly show that the proposed technique can outperform the reference transform domain DVC codec [18] by a significant margin. Potential applications of the proposed DVC codec includes fixed camera applications such as video surveillance systems and PC cameras.

The proposed DVC techniques in this chapter have been published in a journal and a conference (please see Appendix B).

Chapter 7

7 Conclusions

DVC is a recent solution for video coding technology that is built upon the concept of distributed source coding. The objective of the thesis is to improve the DVC paradigm and bring it towards a practical realisation. In order to achieve this goal, this research is focused on a number of areas of the DVC framework:

- Side information refinement
- Non-linear quantisation
- Unidirectional DVC
- Modified frame coding structures

The specific technical contributions presented in this thesis are summarised below.

7.1 Technical Contributions of the Thesis

7.1.1 Iterative Side Information Refinement Techniques

The third chapter few proposes novel side information improvement techniques for transform domain DVC. The nature of the transform coding is exploited and some relevant information is extracted from the decoded DC frame. Side information is refined by the help of this extracted information and used for decoding the remaining coefficient bands. Moreover, another technique which makes use of some other DCT coefficient bands as well as DC coefficients is also given in Section 3.4. Simulation results clearly show that the proposed techniques outperform the state of the art, by a significant margin. Furthermore, results suggest that the proposed codec decreases the performance difference between DVC and H.264/AVC. When the two proposed techniques are compared, the latter (proposed in Section 3.4) further improves the performance since more information is available for the refinement. Also it is obvious that the proposed techniques do not increase the encoder complexity as they are implemented into the DVC decoder.

Another novel side information generation technique based on conventionally intra-coding the DC frame is proposed in Section 3.5. This technique is also built on the DC motion refinement algorithm. At the encoder the DC frame is intra-coded and at the decoder the decoded DC frame is used for generating the side information together with the reference frames. Simulation results clearly show that the proposed codec outperforms the VISNET-II codec [6] by a significant margin.

7.1.2 Residual Coding for Transform Domain DVC

In Section 4.1, a residual coding technique is presented where the encoder exploits both spatial and temporal redundancies by DCT based coding and taking the pixel-wise difference of two consecutive frames. Residual frames are coded as Wyner-Ziv frames in the proposed scheme. Moreover, a novel quantisation technique is proposed to optimise the rate-distortion performance. Simulation results show that there is a significant improvement in the objective quality at the same bit rate compared to reference transform domain Wyner-Ziv codec [18], at the expense of a minor increase in computational complexity of the encoder.

7.1.3 Design of a Non-linear Quantisation Algorithm

In the second part of the fourth chapter, a nonlinear quantisation technique is presented considering the signal characteristics of residual frames. The residual signal of the Wyner-Ziv frame, derived by taking the difference with the preceding key frame, demonstrates a distinct pdf with a very high probability concentration around the near zero mean. Therefore, a non-linear quantisation algorithm is proposed for the DVC codec to exploit the dominating contribution from the relatively small values of the DCT coefficient bands concentrated near zero. The simulation results illustrate a consistent improved RD performance at all bit rates when different test video sequences with varying motion levels are considered.

7.1.4 Block Based Encoder Rate Allocation Algorithm for UDVC

Chapter 5 presents a transform domain UDVC solution. The proposed codec is a practical solution to the main problem of having a feedback channel in DVC, which prevents DVC from being used in a number of consumer electronics applications. At the encoder, parity bits are allocated dynamically based on the motion activity of each block. At the decoder, these parity bits are used to decode bit planes. Simulation results clearly show that the RD performance of the proposed codec can approach that of the DVC codec with feedback, which is assumed to be the upper limit for UDVC. However, it is observed that erroneous DCT coefficients may result in annoying artefacts. Therefore, an improved reconstruction algorithm is adopted in order to increase the

subjective quality. The results clearly show the improvement by using the improved reconstruction technique.

7.1.5 Interlaced Coding Technique for Exploiting Spatial Correlations

Section 6.1 presents the proposed transform domain DVC codec exploiting spatial correlations at the decoder. The proposed technique is an improved solution for especially high motion sequences. Frames are divided into two sub-frames; one is intra-coded using a conventional intra-coding and the other is WZ coded. At the decoder, the prediction of the WZ sub-frame is generated by two different techniques: median prediction and pixel interpolation. The memory requirement at the encoder and decoder is significantly reduced and also the overall latency is reduced since this is an intra-coding technique (no temporal dependency is considered in the proposed algorithm). Simulation results clearly show that the proposed codec outperforms the reference transform domain DVC codec [18] by a significant margin.

7.1.6 Region of Activity Based Coding

In the second part of the sixth chapter, a region-of-activity based technique is proposed. In this technique, Wyner-Ziv frames are divided into macro blocks and then the macro blocks are re-ordered according to the level of motion activity. Motion activity is estimated by simple calculations at the encoder: the SAD of two neighbouring key frames. Therefore, parity bits have been more efficiently utilised yielding a significant performance gain. Simulation results clearly show that the proposed technique can outperform the reference transform domain DVC codec [18] by a significant margin. Potential application of the proposed DVC codec includes fixed camera applications such as video surveillance systems and PC cameras.

7.2 Suggested Future Work

Following areas are identified as potential topics for future research based on experience and the vision achieved through this research and review of peer research:

- **Side Information Generation:**

As discussed in the previous chapters, side information generation has been one of the most popular research subjects. Ideal side information means, there is no need to transmit parity bits for the Wyner-Ziv frame. There have been several improvements on side information generation process, including the side information refinement technique presented in this thesis. However, it

is noted that side information accuracy is still not high enough for some video sequences containing high motion. Therefore, the performance of DVC can not reach that of conventional video coding.

- **Higher GOP sizes:**

In most of the related work, group of pictures size is accepted as two which is not realistic for a simple encoder design. Since the most of the encoding complexity is because of encoding of the key frames. This can be done by defining other types of frames as well as key frames and Wyner-Ziv frames. The optimum performance and encoder complexity can be achieved by exploring spatio-temporal correlations for the side information at higher GOP sizes.

- **Unidirectional DVC:**

As discussed in the fifth chapter, most of the DVC solutions need a feedback channel. However, for many practical applications this is not feasible. Encoder rate allocation techniques are necessary for unidirectional DVC. However, any complex operations are avoided at DVC encoder. As still there is a large performance gap between DVC and unidirectional DVC, this is an active area for researchers.

- **Wyner-Ziv Decoder Complexity:**

There are not many discussions on the complexity of Wyner-Ziv decoder as the most of attention is concentrated on the encoder complexity. For DVC being a practical solution, the decoder complexity has also to be considered.

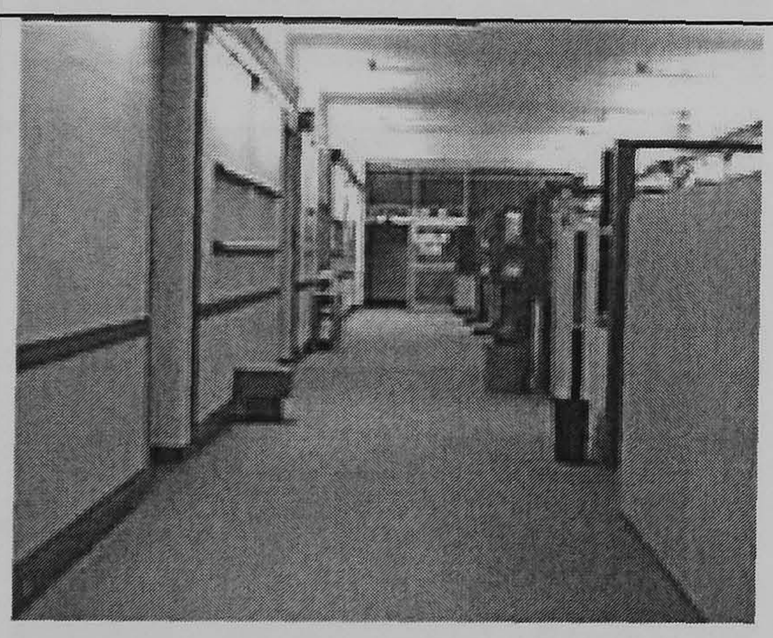
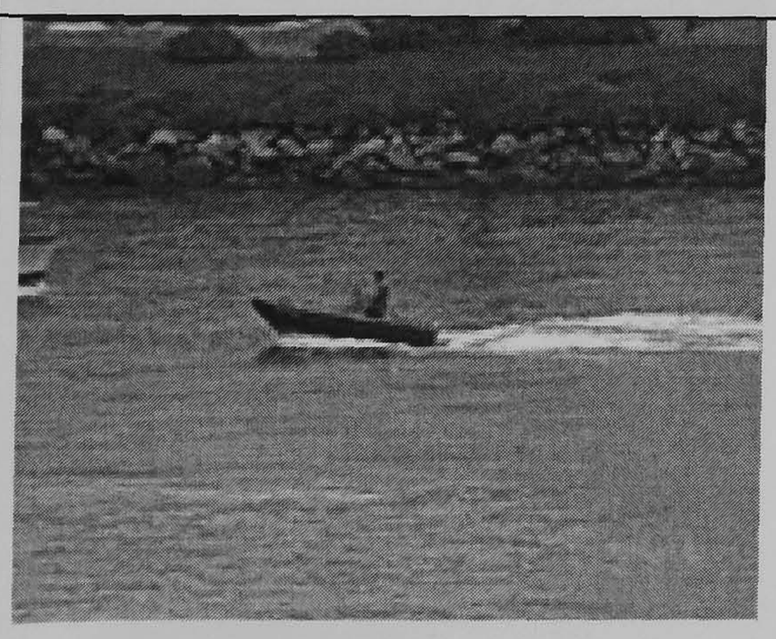
- **Encoding Key Frames:**

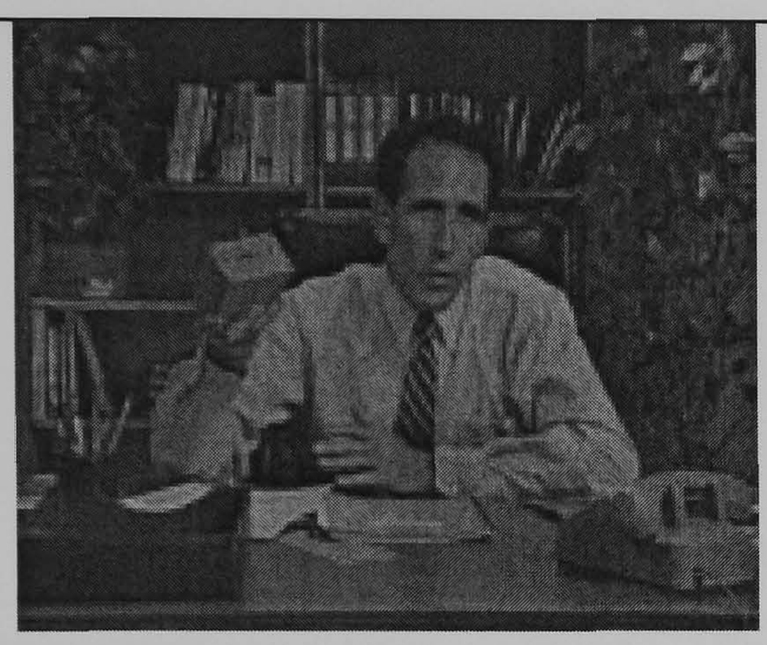
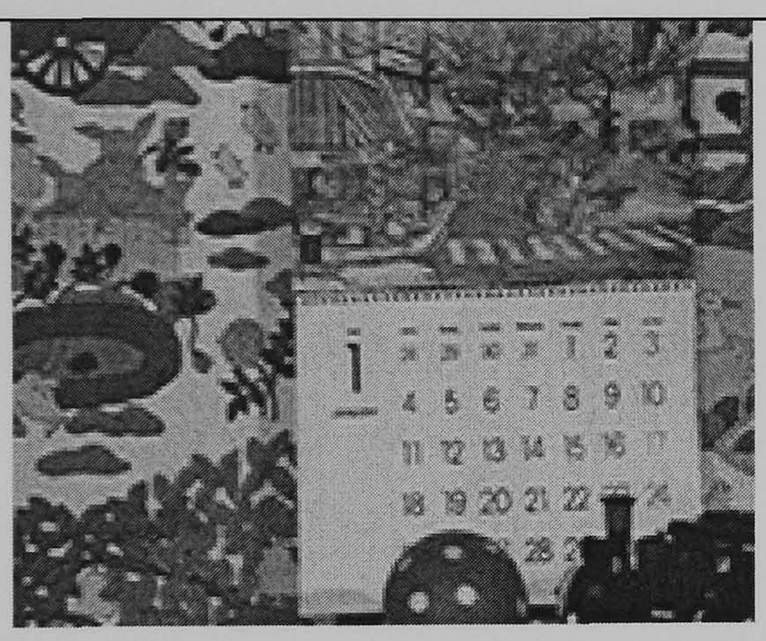
In the current DVC codecs, key frame coding has a very high contribution in the encoder complexity. Therefore DVC encoder complexity can be considerably reduced by reducing the key frame encoding complexity. This can be another research area for the future.

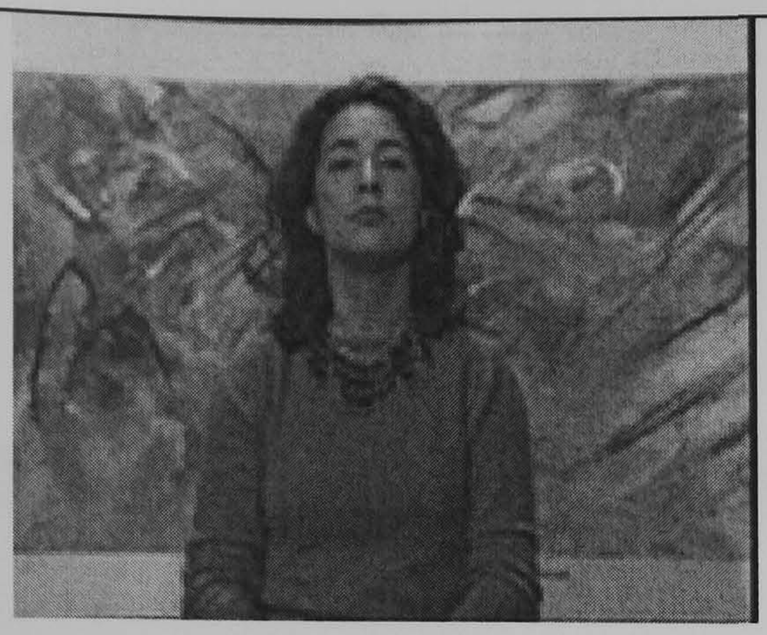
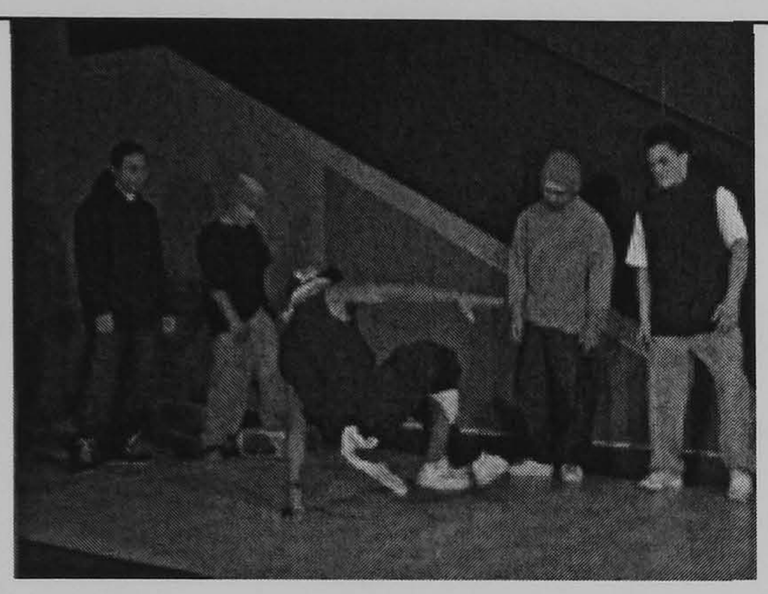
Appendix A

Test Videos

The test video sequences used for simulations in this thesis are listed below with some basic details.

Frame #0	Sequence name	Frame format	Temporal Resolution (fps)	Motion activity level
	Foreman	QCIF (spatial res: 176×144; sampling: 4:2:0)	30, 15	Medium
	Hall-Monitor	QCIF	30, 15	Low
	Coastguard	QCIF	30, 15	Low - medium

	Soccer	QCIF	30, 15	High
	Salesman	QCIF	30, 15	Low
	Mobile	QCIF	30, 15	Medium
	Mother and Daughter	QCIF	30,15	Low

	Silent	QCIF	30,15	Low
	Breakdancers (view 0)	256×192 4:2:0	30,15	High

Appendix B

List of Publications

- M.B. Badem, H.Kodikara Arachchi, S.T.Worrall, A.M.Kondoz, "Transform Domain Residual Coding Technique for Distributed Video coding," *Proceedings of Picture Coding Symposium (PCS)*, Lisbon, Portugal, November 2007.

- M.B. Badem, M. Mrak and W.A.C. Fernando, "Side information refinement using motion estimation in DC domain for transform-based distributed video coding," *IET Electronics Letters*, Vol: 44, No: 16, pp. 965–966, July 2008.

- M.B. Badem, W.A.C. Fernando, W.A.R.J. Weerakkody, S.L.P. Yasakethu and A. M. Kondoz, "Region-of-Activity Based Coding for Transform Domain DVC," *4th International Conference on Information and Automation for Sustainability (ICIAFS)*, Colombo, Sri Lanka, December 2008.

- S.L.P. Yasakethu, W.A.C. Fernando, M.B. Badem and A. M. Kondoz, "Quality Analysis of Side Information Refinement in Transform-Based DVC," *4th International Conference on Information and Automation for Sustainability (ICIAFS)*, Colombo, Sri Lanka, December 2008.

- W.A.R.J. Weerakkody, W.A.C. Fernando, M.B. Badem and A.M. Kondoz, "Nonlinear quantisation for pixel domain distributed video coding," *IET Electronics Letters*, Vol: 45, No: 5, pp. 261–262, February 2009.

- M.B. Badem, W.A.R.J. Weerakkody, W.A.C. Fernando, A.M. Kondoz, "Design of a Non-Linear Quantizer for Transform Domain DVC," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer*, Vol: E92-A, No: 3, pp. 847–852, March 2009.

- M.B. Badem, W.A.C. Fernando, W.A.R.J.Weerakkody, H.Kodikara Arachchi, A.M.Kondoz, "Transform Domain Unidirectional Distributed Video Coding Using Dynamic Parity Allocation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer*, Vol.E92-A, No.4, pp.1202–1208, April 2009.

- M.B. Badem, W.A.C. Fernando, J.L. Martinez and P. Cuenca, "An Iterative Side Information Refinement Technique for Transform Domain Distributed Video Coding," *Proceedings of IEEE International Conference on Multimedia & Expo (ICME)*, New York, USA, June 2009.

- M.B. Badem, W.A.C. Fernando, A.M. Kondoz, "Transform Domain Distributed Video Coding with Spatial Correlations," accepted for publication in *Multimedia Tools and Applications*.
- M.B. Badem, W.A.C. Fernando, M. Mrak, "Improved side information generation for transform domain distributed video coding," submitted to *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2010* in September 2009.

Appendix C

DVC Test Conditions

Key frame quantisation parameters for the DVC codec used in [18], for GOP = 2 at 15 fps:

Sequence	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Foreman	42	40	39	36	35	33	31	26
Hall-Monitor	37	36	35	33	32	31	29	25
Coastguard	39	38	38	35	34	33	31	27
Soccer	45	44	42	38	38	35	31	26

Key frame quantisation parameters for VISNET-II DVC codec (same as DISCOVER DVC test conditions [5]), for GOP = 2 at 15 fps:

Sequence	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Foreman	40	39	38	34	34	32	29	25
Hall-Monitor	37	36	36	33	33	31	29	24
Coastguard	38	37	37	34	33	31	30	26
Soccer	44	43	41	36	36	34	31	25

Bibliography

[1] A. Aaron, R. Zhang and B. Girod, "Wyner-Ziv coding of motion video," *Proceedings of Asilomar Conference on Signals and Systems*, Pacific Grove, CA, November 2002.

[2] J. Q. Pedro, L. D. Soares, C. Brites, J. Ascenso, F. Pereira, C. Bandeirinha, S. Ye, F. Dufaux, T. Ebrahimi, "Studying Error Resilience Performance for a Feedback Channel Based Transform Domain Wyner-Ziv Video Codec", *Picture Coding Symposium (PCS)*, Lisbon, Portugal, November 2007.

[3] DISCOVER, "Application scenarios and functionalities for DVC," Tech. Report Deliverable 19, August 2007.

[4] R. Puri, K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," *Proceedings of 40th Allerton Conference on Communication, Control, and Computing*, Allerton, IL, October 2002.

[5] http://www.discoverdvc.org/

[6] http://www.visnet-noe.org/

[7] M. Ghanbari, "Standard Codecs: Image Compression to Advanced Video Coding," The Institution of Electrical Engineers, London, UK, 2003.

[8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", *IEEE Transactions on Image Processing*, Vol: 13, pp. 600–612, April 2004.

[9] A. P. Hekstra, J. G. Beerends, D. Ledermann, F. E. de Caluwe, S. Kohler, R. H. Koenen, S. Rihs, M. Ehrsam and D. Schlauss, "PVQM- a perceptual video quality measure" *Signal Processing: Image Communication*, Vol: 17, No: 10, pp. 781–798, November 2002.

[10] ITS Video quality research, www.its.bldrdoc.gov/n3/video/VQM_software.php, 2007.

[11] ISO/IEC International Standard 14496-10:2003, "Information Technology – Coding of Audio-visual Objects – Part 10: Advanced Video Coding".

[12] J.D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Inform Theory*, Vol: IT-19, pp. 471–480, July 1973.

[13] A.D. Wyner, J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, Vol: IT-22, pp. 1–10, January 1976.

[14] T. Wiegand, G. Sullivan, G. Bjøntegaard and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol: 13, No: 7, pp. 560–576, July 2003.

[15] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-Domain Wyner-Ziv Codec for Video," *Proceedings of VCIP*, San Jose, USA, January 2004.

[16] A. Aaron, S. Rane and B. Girod, "Wyner-Ziv Video Coding with Hash-Based Motion Compensation at the Receiver," *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Vol: 5, pp. 3097–3100, Singapore, October 2004.

[17] J. Ascenso, C. Brites and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," *Proceedings of 5th EURASIP Conference on Speech and Image Processing, Multimedia, Communications and Services (EUSIPCO)*, Smolenice, Slovak, July 2005.

[18] C. Brites, J. Ascenso, and F. Pereira, "Improving Transform Domain Wyner-Ziv Video Coding Performance," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France, May 2006.

[19] A. Aaron, E. Setton and B. Girod, "Towards practical Wyner-ziv coding of video," *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Barcelona, Spain, pp. 869–872, September 2003.

[20] S. Pradhan and K. Ramchandran, "DIstributed Source Coding Using Syndromes (DISCUS): Design and construction," *Proceedings of IEEE Data Compression Conference (DCC)*, pp. 158–167, Snowbird, (Utah), USA, March 1999.

[21] X. Wang and M. Orchard, "Design of trellis codes for source coding with side information at the decoder," *Proceedings of IEEE Data Compression Conference*, pp. 361–370, Snowbird, Utah, USA, March 2001.

[22] J. Garcia-Frias, "Compression of correlated binary sources using turbo codes," *IEEE Communications Letters*, Vol: 5, No: 10, pp. 417–419, October 2001.

[23] J. Garcia-Frias and Y. Zhao, "Data compression of unknown single and correlated binary sources using punctured turbo codes," *Proceedings of 39th Allerton Conference on Communication, Control and Computing*, Monticello, (Illinois), USA, October 2001.

[24] J. Bajcsy and P. Mitran, "Coding for the Slepian-Wolf Problem with Turbo Codes," *Proceedings of IEEE Global Telecommunications Conference*, Vol: 2, pp.1400–1404, San Antonio, Texas, USA, November 2001.

[25] P. Mitran and J. Bajcsy, "Near Shannon-Limit Coding for the Slepian-Wolf Problem," *Proceedings of 21st Biennal Symposium on Communications*, Kingston, Ontario, Canada, June 2002.

[26] A. Aaron and B. Girod, "Compression with side information using turbo codes," *Proceedings of IEEE Data Compression Conference (DCC)*, pp. 252–261, Snowbird, (Utah), USA, April 2002.

[27] A. Liveris, Z. Xiong and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, Vol: 6, No: 10, pp. 440–442, October 2002.

[28] A. Liveris, Z. Xiong and C. Georghiades, "Compression of binary sources with side information using low-density parity-check codes," *Proceedings of IEEE Global Telecommunications Conference*, pp. 1300–1304.Taipei, Taiwan, November 2002.

[29] A. Liveris, Z. Xiong and C. Georghiades, "Distributed Compression of Binary Sources Using Conventional Parallel and Serial Concatenated Convolutional Codes," *Proceedings of IEEE Data Compression Conference*, pp.193–202, Snowbird, Utah, USA, March 2003.

[30] V. Stankovic, A. Liveris, Z. Xiong and C. Georgiades, "Design of Slepian-Wolf Codes by Channel Code Partitioning," *Proceedings of IEEE Data Compression Conference*, pp. 302–311, Snowbird, Utah, USA, March 2004.

[31] D. Schonberg, S. Pradhan, and K. Ramchandran, "Distributed Code Constructions for the Entire Slepian-Wolf Rate Region for Arbitrarily Correlated Sources," *Proceedings of IEEE Data Compression Conference*, pp. 292–301, Snowbird, Utah, USA, March 2004.

[32] T. Coleman A. Lee, M. Medard and M. Effros, "On Some New Approaches to Practical Slepian-Wolf Compression Inspired by Channel Coding," *Proceedings of IEEE Data Compression Conference*, pp. 282–291, Snowbird, Utah, USA, March 2004.

[33] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes," *Proceedings of IEEE International Conference on Communications (ICC)*, Geneva, Switzerland, pp. 1064–1070, May 1993.

[34] W.A.R.J. Weerakkody, "Towards Practical Distributed Video Coding," Phd. Thesis, University of Surrey, UK, September 2007.

[35] M. Tagliasacchi and S. Tubaro, "A MCTF Video Coding Scheme Based on Distributed Source Coding Principles," *Visual Communication and Image Processing*, July 2005.

[36] L. Natario, C. Brites, J. Ascenso and F. Pereira, "Extrapolating Side Information for Low-Delay Pixel Domain Distributed Video Coding," *International Workshop on Very Low Bitrate Video Coding*, Sardinia, Italy, September 2005.

[37] J. Ascenso, C. Brites and F. Pereira, "Motion Compensated Refinement for Low Complexity Pixel Based Distributed Video Coding," *Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Como, Italy, July 2005.

[38] L. Alparone, M. Barni, F. Bartolini and V. Cappellini, "Adaptively Weighted Vector-Median Filters for Motion-Fields Smoothing," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 4, pp. 2267–2270, Georgia, USA, May 1996.

[39] C. Brites, J. Ascenso, and F. Pereira, 'Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding," *Proceedings of IEEE International Conference on Image Processing, (ICIP)*, Atlanta, CA, USA, October 2006.

[40] Meyer, Westerlaken, Gunnewiek, Lagendijk, "Distributed Source Coding of Video with Non-Stationary Side Information," *Proceedings of SPIE VCIP*, Beijing, China, June 2005.

[41] M. Tagliasacchi, J. Pedro, F. Pereira, S. Tubaro, "An Efficient Request Stopping Method at the Turbo Decoder in Distributed Video Coding," *EURASIP European Signal Processing Conference (EUSIPCO 2007)* , September 2007.

[42] C. Brites, "Advances on Distritubed Video Coding," Master's thesis, Instituto Superior Tecnico, Lisbon, Portugal, December 2005.

[43] K. R. Rao and P. Yip, "Discrete Cosine Transform: Algorithms, Advantages, Applications," Academic Press, Boston, USA, 1990.

[44] X. Guo, Y. Lu, F. Wu, W. Gao, "Distributed Video Coding Using Wavelet," *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 5427–5430, May 2006.

[45] D. Kubasov, C. Guillemot, "Mesh-Based Motion-Compensated Interpolation for Side Information Extraction in Distributed Video Coding," *Proceedings of IEEE International Conference on Image Processing, (ICIP)*, pp 261 – 264, Atlanta, USA, October 2006.

[46] T.N. Dinh, G. Lee, J.Y. Chang, H. Cho, "Side Information Generation Using Extra Information in Distributed Video Coding," *Proceedings of IEEE International Symposium on Signal Processing and Information Technology*, pp. 138 – 143, Dec 2007.

[47] W. Chien, L.J. Karam, G.P. Abousleman, "Distributed Video Coding With Lossy Side Information," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vol: 2, Toulouse, France, May 2006.

[48] Lei Wei, Yao Zhao, Anhong Wang, "Improved Side-Information in Distributed Video Coding," *Proceedings of First International Conference on Innovative Computing, Information and Control (ICICIC)*, Vol: 2, pp. 209 – 212, August 2006.

[49] M. Tagliasacchi, S. Tubaro, A. Sarti, "On the Modeling of Motion in Wyner-Ziv Video Coding," *IEEE International Conference on Image Processing (ICIP)*, pp. 593–596, October 2006.

[50] M. Dalai, R. Leonardi, F. Pereira, "Improving Turbo Codec Integration In Pixel-Domain Distributed Video Coding," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vol: 3, pp. III-257–III-260, Toulouse, France, May 2006.

[51] T.N. Dinh, G. Lee, J. Chang and H. Cho, "A Novel Motion Compensated Frame Interpolation Method for Improving Side Information in Distributed Video Coding," *International Symposium on Information Technology Convergence (ISITC)*, pp 179–183, November 2007.

[52] M. Guo, Y. Lu, F. Wu, S. Li and W. Gao, "Distributed Video Coding with Spatial Correlation Exploited Only at the Decoder," *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 41–44, May 2007.

[53] Y. Zhang, C. Zhu, "Full Search of Side-Information in Distributed Video Coding," *International Conference on Image and Graphics (ICIG)*, pp. 246–249, August 2007.

[54] M. Tagliasacchi, L. Frigerio, S. Tubaro, "Rate-Distortion Analysis of Motion-Compensated Interpolation at the Decoder in Distributed Video Coding," *IEEE Signal Processing Letters*, Vol: 14, Issue: 9, pp. 625–628, September 2007.

[55] Y. Vatis, S. Klomp, J. Ostermann, "Enhanced reconstruction of the quantized transform coefficients for Wyner-Ziv coding", *Proceedings of IEEE International Conference on Multimedia & Expo (ICME)*, Beijing, China, July 2007.

[56] D. Kubasov, J. Nayak, C. Guillemot, "Optimal Reconstruction in Wyner-Ziv Video Coding with Multiple Side Information," *Proceedings of International Workshop on MMSP*, Greece, October 2007.

[57] W.A.R.J. Weerakkody, W.A.C. Fernando, A.M.Kondoz, "An Enhanced Reconstruction Algorithm for Unidirectional Distributed Video Coding," *12th IEEE International Symposium on Consumer Electronics (ISCE)*, Algarve, Portugal, April 2008.

[58] Y. Tonomura, T. Nakachi, "A New Framework for Distributed Video Coding Based on JPEG 2000," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vol: 3, May 2006.

[59] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, M. Ouaret, "The DISCOVER codec: Architecture, Techniques and Evaluation", *Proceedings of Picture Coding Symposium (PCS)*, Lisbon, Portugal, November 2007.

[60] M. Ouaret, F. Dufaux and T. Ebrahimi, "On Comparing Image and Video Compression Algorithms", *International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, Scottsdale, Arizona, U.S.A, January 2007.

[61] M. Maitre, C. Guillemot, L. Morin, "3D Scene Modeling for Distributed Video Coding," *Proceedings of IEEE International Conference on Image Processing (ICIP)*, pp. 585–588, October 2006.

[62] X. Artigas, E. Angeli, L. Torres, "Side Information Generation for Multiview Distributed Video Coding Using a Fusion Approach," *Proceedings of the 7th Nordic Signal Processing Symposium (NORSIG)*, pp. 250–253, June 2006.

[63] C. Tonoli, P. Migliorati and R. Leonardi, "Error Resilience in Current Distributed Video Coding Architectures," *EURASIP Journal on Image and Video Processing*, Vol: 2009, Article ID 946585, 2009.

[64] C. Tonoli, M. Dalai, P. Migliorati, R. Leonardi, "Error Resilience Performance Evaluation of A Distributed Video Codec," *Proceedings of Picture Coding Symposium*, Lisbon, Portugal, November 2007.

[65] X. Artigas, L. Torres, "Iterative generation of motion-compensated side information for distributed video coding", *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Genova, Italy, September 2005.

[66] A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, H.K. Arachchi, "A Sequential Motion Estimation Using Luminance And Chrominance Information For Distributed Video Coding Of Wyner-Ziv Frames," *IEE Electronics Letters*, Vol: 42, No: 7, pp 396–397, March 2006.

[67] A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, "Multiple Side Information Streams for Distributed Video Coding," *IET electronics Letters*, Vol: 42, No: 25, pp: 1447–1449, December 2006.

[68] A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, "Side Information Improvement in DVC with Two Side Information Streams and 3D Motion Refinement," *Proceedings of IEEE CCECE*, Vancouver, Canada, April 2007.

[69] W.A.R.J. Weerakkody, W.A.C. Fernando, J.L. Martínez, P.Cuenca, F.Quiles, "An Iterative Refinement Technique for Side Information Generation in DVC," *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, Beijing, China, July 2007.

[70] M.B. Badem, M. Mrak and W.A.C. Fernando, "Side information refinement using motion estimation in DC domain for transform-based distributed video coding," *IET Electronics Letters*, Vol. 44, No. 16, pp. 965 – 966, July 2008.

[71] M.B. Badem, W.A.C. Fernando, J.L. Martinez and P. Cuenca, "An Iterative Side Information Refinement Technique for Transform Domain Distributed Video Coding," *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, New York, USA, June 2009.

[72] S. Ye, M. Ouaret, F. Dufaux, and T. Ebrahimi, "Improved Side Information Generation for Distributed Video Coding by Exploiting Spatial and Temporal Correlations," *EURASIP Journal on Image and Video Processing*, 2009.

[73] A. Aaron, D. Varodayan and B. Girod, "Wyner-Ziv residual coding of video," *Proceedings of Picture Coding Symposiu (PCS)*, Beijing, China, April 2006.

[74] M.B. Badem, H.Kodikara Arachchi, S.T.Worrall, A.M.Kondoz, "Transform Domain Residual Coding Technique for Distributed Video coding," *Proceedings of Picture Coding Symposium (PCS)*, Lisbon, Portugal, November 2007.

[75] X. Artigas and L. Torres, "Improved signal reconstruction and return channel suppression in Distributed Video Coding systems," *Proceedings of 47th International Symposium ELMAR: Multimedia Systems and Applications*, Zadar, Croatia, June 2005.

[76] A.B.B. Adikari, W.A.C. Fernando, W.A.R.J. Weerakkody, "Iterative Wyner-Ziv decoding for unidirectional distributed video coding," *IEE Electronics Letters*, Vol: 43, No: 2, pp. 93–95. January 2007.

[77] W.A.R.J. Weerakkody, W.A.C. Fernando and A.B.B. Adikari, "Unidirectional Distributed Video Coding for Low Cost Video Encoding," *IEEE Transactions on Consumer Electronics*, Vol: 53, Issue: 2, pp. 788–795, May 2007.

[78] M. Morbee, J. Prades-Nebot, A. Pizurica and W. Philips, "Rate allocation algorithm for pixel-domain distributed video coding without feedback channel," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, (Hawaii), USA, pp. I-521–I-524, April 2007.

[79] M. Morbee, J. Prades-Nebot, A. Roca, A. Pizurica and W. Philips, "Improved pixel-based rate allocation for pixel-domain distributed video coders without feedback channel *Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS)*, Delft, The Netherlands, pp. 663–674, August 2007.

[80] C. Brites, F. Pereira, "Encoder Rate Control for Transform Domain Wyner-Ziv Video Coding," *Proceedings of IEEE International Conference on Image Processing (ICIP)*, San Antonio, Texas, USA, September 2007.

[81] J.L. Martinez, C. Holder, G. Fernandez-Escribano, H. Kalva and F. Quiles, "DVC using a half-feedback based approach," *Proceedings of International Conference on Multimedia and Expo (ICME)*, Hannover, Germany, pp. 1125–1128, June 2008.

[82] J.L. Martinez, G. Fernandez-Escribano, H. Kalva, A. Garrido, W.A.R.J. Weerakkody and W.A.C. Fernando, "Feedback free DVC architecture using machine learning," *Proceedings of IEEE International Conference on Image Processing (ICIP)*, San Diego, (California), USA, pp. 1140–1143, October 2008.

[83] M. Dalai, R. Leonardi, F. Pereira, "Improving Turbo Codec Integration In Pixel-Domain Distributed Video Coding," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vol: 3, pp. 257–260, May 2006.

[84] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: turbo-codes," *Proceedings of IEEE International Conference on Communications*, Geneva, Switzerland, pp. 1064–1070, May 1993.

[85] Y. Vatis, S. Klomp, J. Ostermann, "Enhanced reconstruction of the quantized transform coefficients for Wyner-Ziv coding", *Proceedings of IEEE International Conference on Multimedia & Expo (ICME)*, Beijing, China, July 2007.

[86] D. Kubasov, J. Nayak, C. Guillemot, "Optimal Reconstruction in Wyner-Ziv Video Coding with Multiple Side Information," *Proceedings of International Workshop on MMSP*, Greece, October 2007.