Intra-Key-Frame Coding and Side Information Generation Schemes in Distributed Video Coding

Suvendu Rup

Department of Computer Science and Engineering

National Institute of Technology Rourkela

Rourkela – 769 008, India

Intra-Key-Frame Coding and Side Information Generation Schemes in Distributed Video Coding

Dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Computer Science and Engineering

by

Suvendu Rup

(Roll- 508CS103)

under the guidance of

Prof. Banshidhar Majhi

Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela, Odisha, 769 008, India

October 2013

dedicated to my Parents...

Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela-769 008, Odisha, India.

Dr. Banshidhar Majhi
Professor

November 6, 2013

Certificate

This is to certify that the work in the thesis entitled ***Intra-Key-Frame Coding and Side Information Generation Schemes in Distributed Video Coding*** by ***Suvendu Rup*** is a record of an original research work carried out by him under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy in Computer Science and Engineering in the department of Computer Science and Engineering, National Institute of Technology Rourkela. Neither this thesis nor any part of it has been submitted for any degree or academic award elsewhere.

Banshidhar Majhi

Acknowledgement

This dissertation, though an individual work, has benefited in various ways from several people. Whilst it would be simple to name them all, it would not be easy to thank them enough.

The enthusiastic guidance and support of *Prof. Banshidhar Majhi* inspired me to stretch beyond my limits. His profound insight has guided my thinking to improve the final product. My solemnest gratefulness to him.

I am also grateful to *Prof. Ratnakar Dash* for his ceaseless support throughout my research work. My sincere thanks to *Prof. Pankaj Kumar Sa* for his continuous encouragement and invaluable advice.

It is indeed a privilege to be associated with people like *Prof. S. K. Rath, Prof. S.Padhy, Prof. S.K.Jena, Prof. Gagan Rath, Prof. D. P. Mohapatra, Prof. A. K. Turuk, Prof. S.Chinara, Prof. B. D. Sahoo* and *Prof. P. M. Khilar*. They have made available their support in a number of ways.

Many thanks to my comrades and fellow research colleagues. It gives me a sense of happiness to be with you all. Special thanks to *Hunny, Saroj, Ashis, Anshuman, Soubhagya* whose involvement gave a new breadth to my research.

Finally, my heartfelt thanks to my wife *Sukirti* and son *Samprit* for their unconditional love and support. Words fail to express my gratitude to my beloved parents who sacrificed their comfort for my betterment.

Suvendu Rup

Abstract

In this thesis investigation has been made to propose improved schemes for intra-key-frame coding and side information (SI) generation in a distributed video coding (DVC) framework. From the DVC developments in last few years it has been observed that schemes put more thrust on intra-frame coding and better quality side information (SI) generation. In fact both are interrelated as SI generation is dependent on decoded key frame quality. Hence superior quality key frames generated through intra-key frame coding will in turn are utilized to generate good quality SI frames. As a result, DVC needs less number of parity bits to reconstruct the WZ frames at the decoder. Keeping this in mind, we have proposed two schemes for intra-key frame coding namely,

(a) Borrows Wheeler Transform based H.264/AVC (Intra) intra-frame coding (BWT-H.264/AVC(Intra))

(b) Dictionary based H.264/AVC (Intra) intra-frame coding using orthogonal matching pursuit (DBOMP-H.264/AVC (Intra))

BWT-H.264/AVC (Intra) scheme is a modified version of H.264/AVC (Intra) scheme where a regularized bit stream is generated prior to compression. This scheme results in higher compression efficiency as well as high quality decoded key frames. **DBOMP-H.264/AVC (Intra)** scheme is based on an adaptive dictionary and H.264/AVC (Intra) intra-frame coding. The traditional transform is replaced with a dictionary trained with K-singular value decomposition (K-SVD) algorithm. The dictionary elements are coded using orthogonal matching pursuit (OMP).

Further, two side information generation schemes have been suggested namely,

(a) Multilayer Perceptron based side information generation (MLP - SI)

(b) Multivariable support vector regression based side information generation (MSVR-SI)

MLP-SI scheme utilizes a multilayer perceptron (MLP) to estimate SI frames from the decoded key frames block-by-block. The network is trained offline using

training patterns from different frames collected from standard video sequences. MSVR-SI scheme uses an optimized multi variable support vector regression (M-SVR) to generate SI frames from decoded key frames block-by-block. Like MLP, the training for M-SVR is made offline with known training patterns apriori.

Both intra-key-frame coding and SI generation schemes are embedded in the Stanford based DVC architecture and studied individually to compare performances with their competitive schemes. Visual as well as quantitative evaluations have been made to show the efficacy of the schemes. To exploit the usefulness of intra-frame coding schemes in SI generation, four hybrid schemes have been formulated by combining the aforesaid suggested schemes as follows:

(a) **BWT-MLP** scheme that uses **BWT-H.264/AVC (Intra)** intra-frame coding scheme and **MLP-SI** side information generation scheme.

(b) **BWT-MSVR** scheme, where we utilize **BWT-H.264/AVC (Intra)** for intra-frame coding followed by **MSVR-SI** based side information generation.

(c) **DBOMP-MLP** scheme is an outcome of putting **DBOMP-H.264/AVC (Intra)** intra-frame coding and **MLP-SI** side information generation schemes.

(d) **DBOMP-MSVR** scheme deals with **DBOMP-H.264/AVC (Intra)** intra-frame coding and **MSVR-SI** side information generation together.

The hybrid schemes are also incorporated into the Stanford based DVC architecture and simulation has been carried out on standard video sequences. The performance analysis with respect to overall rate distortion, number requests per SI frame, temporal evaluation, and decoding time requirement has been made to derive an overall conclusion.

Keywords: Distributed video coding, orthogonal matching pursuit, multilayer perceptron, support vector regression, intra-frame coding, side information, temporal evaluation, rate distortion, peak signal to noise ratio.

Contents

List of Figures

List of Tables

+

Chapter 1

Introduction

Delivering digital video through the available networks needs compression. Compressing video is all about making the best compromises possible without giving up too much quality. The amount of digital contents grows at a rapid rate, so does the demand for communicating them. However, the amount of storage and bandwidth increases at much slower rate. Thus powerful and efficient compression methods plays a crucial role for video storage and transmission.

The current digital video compression schemes are represented by the International Telecommunication Union-Telecommunication (ITU-T) and Motion Picture Expert Group (MPEG) standards, which rely on a combination of block-based transform and inter-frame predictive coding to exploit the spatial and temporal redundancies within the encoded video [1]. This results in high complexity encoder due to the motion estimation task at the encoder side. On the other hand, the decoder is so simple and around five to ten times less complex than the encoder. However, this type of structure is well-suited for applications where the video is encoded once and decoded many times. It shows a one-to-many topologies for down link model applications such as broadcasting or streaming and video on demand.

In recent years, the emerging applications like mobile camera phone, video surveillance, multimedia sensor networks, wireless camera etc. where the memory requirement and computation at the encoder are scarce. The traditional video coding architecture has complex encoder and a simple decoder. It is a challenge for traditional video coding to fulfill the requirements of the above mentioned

applications. Another important goal is to achieve the coding efficiency similar to that of traditional video coding schemes i.e. shifting the complexity from encoder to the decoder should not compromise the coding efficiency [2, 3].

So, to address the challenges, distributed source coding has emerged by exploiting the source statistics partially or totally at the decoder. Thus, it enables a flexible complexity distribution between encoder and decoder [4]. To apply DSC to a video compression standard is called distributed video coding (DVC), targeting both low complexity encoding and error resilience [5, 6].

In this thesis, we address the issues related to DVC and propose four different DVC schemes in subsequent chapters. The present chapter is dedicated to a detail elaboration of fundamentals and recent advances in DVC for complete understanding of motivation behind our work and the suggested schemes. The rest of the chapter is organized as follows. The information theoretic background followed by some promising solutions of DVC are presented in Section 1.1. Section 1.2 reviews the Stanford based architecture followed by Instituto Superior Tecnico (IST) based architecture in detail. Section 1.3 elaborates the other advances in DVC used in Stanford architecture. The research motivation and objectives are formally stated in Section 1.4. Finally, Section 1.5 outlines the layout of the thesis.

1.1   Information theoretic background

Distributed source coding (DSC) theory refers to the coding of two or more statistically dependent random sequences in a distributed way. The term distributed refers to the encoding operation and not to its location. An independent bit stream from various sources are sent from the different encoders and all the encoded bit streams are sent to a single decoder, which performs a joint decoding of all received bit streams. Based on the DSC principle a new video coding paradigm has emerged called DVC with independent encoding and joint decoding principle [7].

Figure 1.1 shows an application scenario of DSC where, multiple dependent camera are sensing the same scene from different position. However, each camera

sends an independent bit stream to a centralized decoder which performs joint decoding of all the received bit streams exploiting the correlation between them. Therefore, it is possible to reduce the complexity of the encoding process by exploiting the correlation between the multiple encoded sequences just at the decoder.
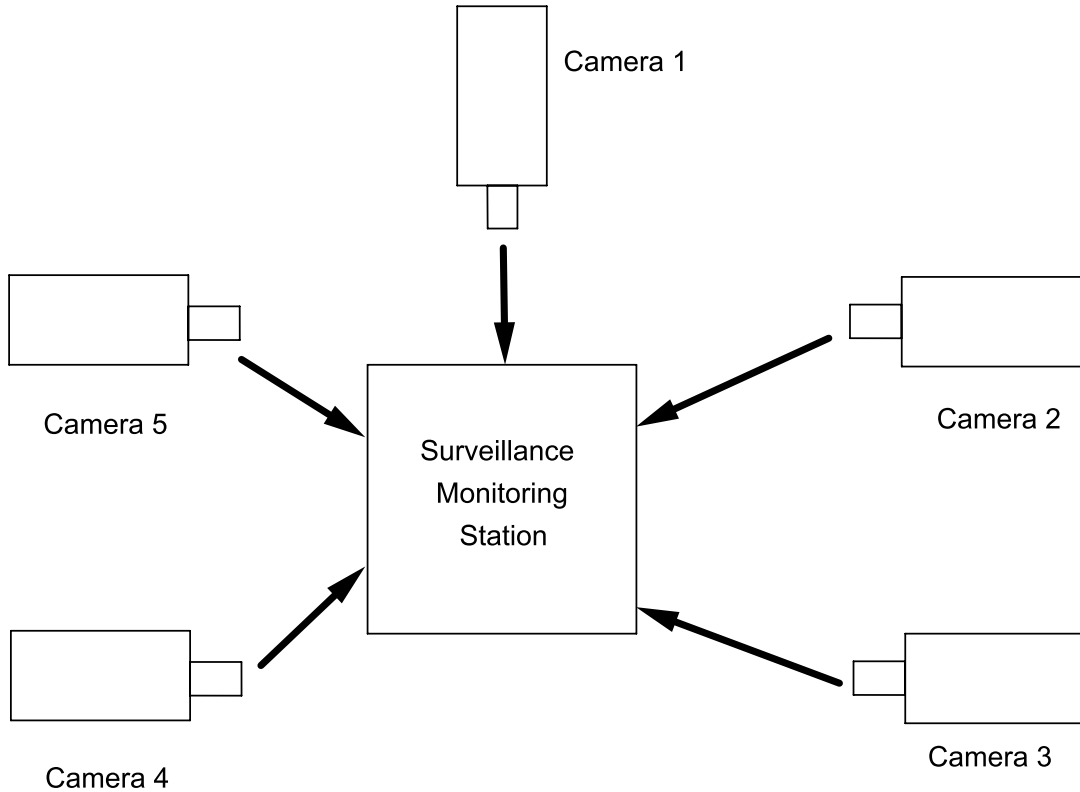
Figure 1.1: A distributed source coding scenario with multiple encoders and a centralized decoder

To understand the traditional video coding from the information theoretic point of view, let us consider two statistically dependent sequences X and Y. As per Shannon's theorem, compression of an independent identically distributed (i.i.d) finite alphabet random sequence source X can be reconstructed without any loss when the rate $R(X)$ is greater than or equal to its entropy $H(X)$.

Figure 1.2 depicts, two statistically dependent sequences X and Y are independently encoded and independently decoded. In this case, the outcome suggests, both sequences are losslessly reconstructed only if

$$R(X) \geq H(X) \tag{1.1}$$

3

and

$$R(Y) \geq H(Y) \tag{1.2}$$
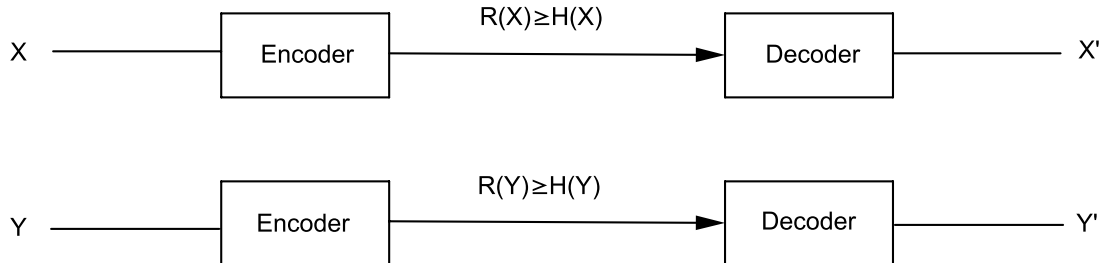
Figure 1.2: Architecture of independent encoding and independent decoding

Figure 1.3 presents, two dependent sequences X and Y are jointly encoded and jointly decoded. The perfect reconstruction of X and Y can be possible if the total rate R is greater than or equal to joint entropy, $H(X, Y)$ as follows,

$$R = R_x + R_y \geq H(X, Y) \tag{1.3}$$
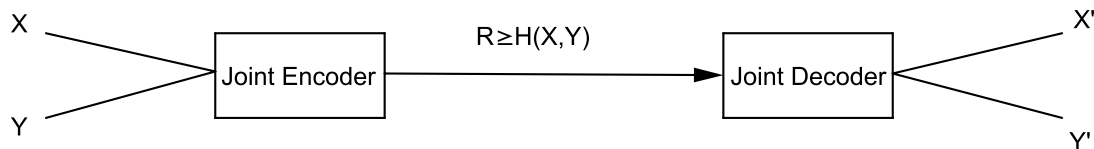
Figure 1.3: Architecture of joint encoding and joint decoding

In Figure 1.4, consider two dependent sequences X and Y are independently encoded and jointly decoded. This scenario presents the perfect architecture of DSC. The possible rate combination R for perfect reconstruction of X and Y is as follows,

$$R \geq H(X) + H(Y) \geq H(X, Y) \tag{1.4}$$

The information theoretic background of DVC is based on two theorems namely Slepian-Wolf theorem and Wyner-Ziv theorem. Both the theorems are discussed in detail.
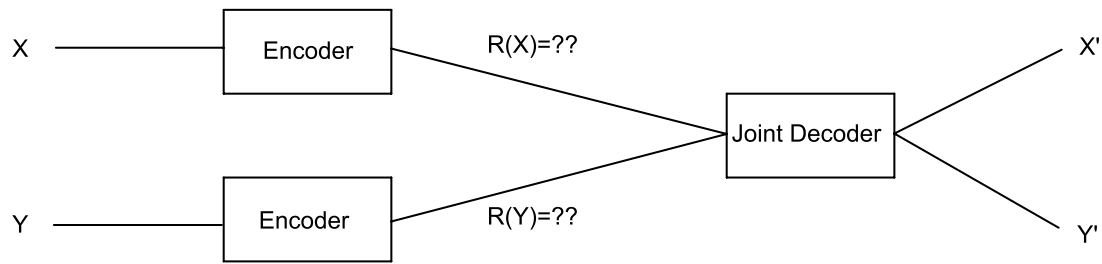
Figure 1.4: Architecture of independent encoding and joint decoding

1.1.1 Slepian Wolf theorem

DVC is a practical realization of the DSC principles that is originally introduced by Slepian and Wolf in 1973. The Slepian-Wolf (SW) theorem suggests the concept of independent encoding and joint decoding [8].

Let us assume X and Y are two independent identically distributed (i.i.d) discrete random sequences. Consider the two sequences are independently encoded with bit rate R_X and R_Y respectively. In 1970, Slepian and Wolf studied this problem and presented an analysis of the possible rate combinations R_X and R_Y for reconstruction of X and Y with an arbitrarily small error probability which can be expressed as,

$$R_X \geq H(X|Y) \tag{1.5}$$

$$R_Y \geq H(Y|X) \tag{1.6}$$

$$R_X + R_Y \geq H(X,Y) \tag{1.7}$$

where, $H(X|Y)$ and $H(Y|X)$ are conditional entropies and $H(X,Y)$ is the joint entropy. Equation (1.7) shows that, despite the separate encoding of X and Y, SW theorem proves that the total rate, $R = R_X + R_Y$, equal to the joint entropy. Therefore, it can be concluded that there should not be loss of compression efficiency theoretically, due to the utilization of independent encoding compared to joint decoding adopted in the traditional video coding standards.

Figure 1.5 shows the achievable rate combination of R_X and R_Y. In Figure 1.5 the vertical, horizontal and diagonal lines represent the equations (1.1), (1.2) and (1.3) respectively.
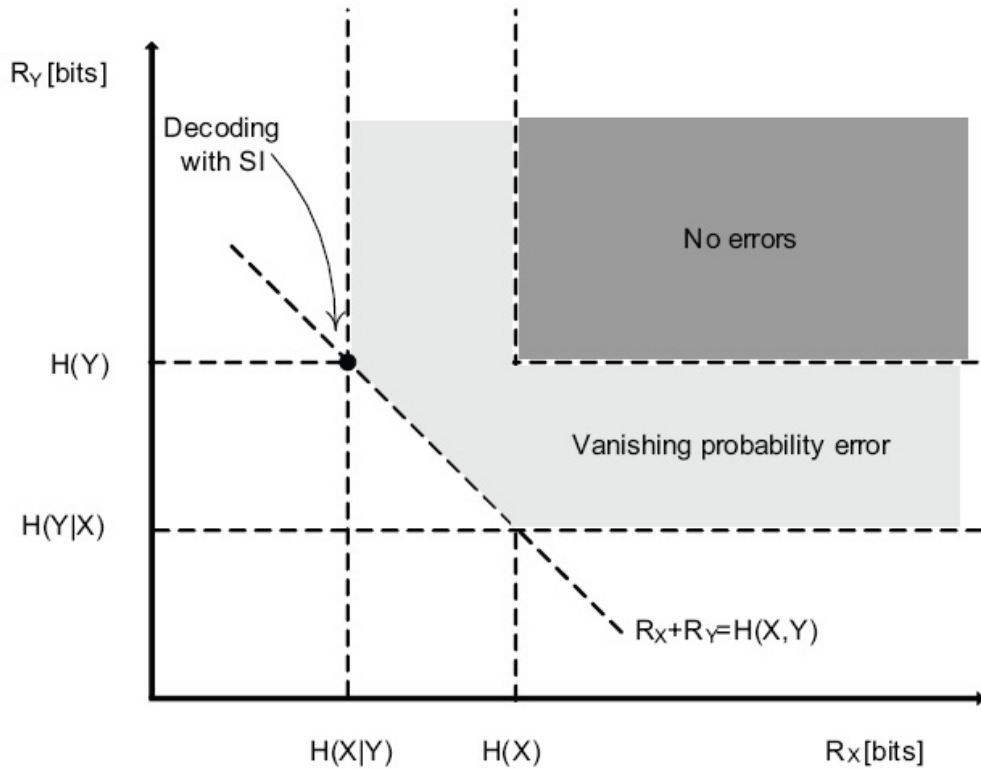
Figure 1.5: Achievable rate region defined by Slepian-Wolf theorem

Another interesting feature of SW coding is the study of channel coding. This feature is studied by Wyner in 1970 [9]. Consider two independent identically distributed binary sequences X and Y and a virtual correlation channel where, the source sequence X known as main information and a channel noisy version of X i.e. Y known as side information (SI). To correct the error between X and Y, a suitable channel code is applied to the sequence X. Then a systematic channel code is applied to encode X and resulting parity bits are transmitted. At the decoder side, the received parity bits and SI (Y) are used to perform error correction for the perfect decoding of X.

1.1.2   Wyner-Ziv theorem

Wyner-Ziv (WZ) theorem is an extension of SW theorem for a lossy case [10]. Consider two i.i.d sequences X and Y, where X is encoded at the encoder end and Y denotes the SI available at the decoder. Figure 1.6 shows the realization of WZ theorem, that suggests the lossy compression with decoder SI.
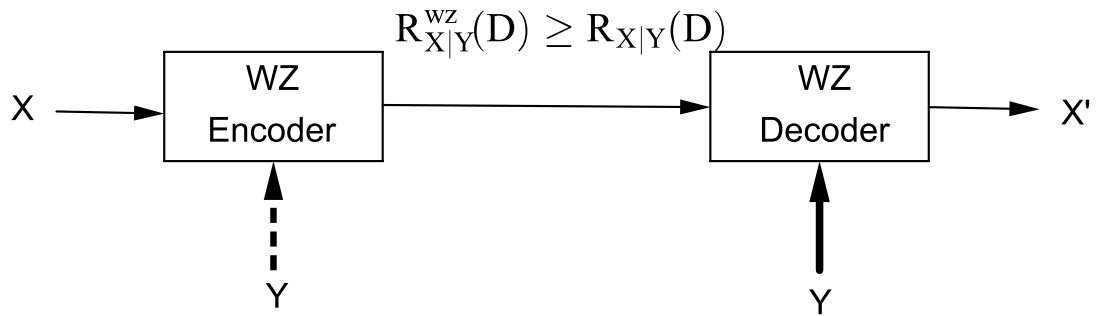
Figure 1.6: Lossy DSC of source X, which depends on SI (Y)

The statistical dependency between X and Y is unknown at the encoder. So the source information X and the corresponding decoded output X' can be obtained at the decoder jointly with received parity bits and SI (Y). Wyner and Ziv have attempted to quantify the minimum bit rate to be transmitted from encoder to the decoder called WZ bit rate $R^{WZ}(D)$, for achieving the finite distortion D between input and output. According to WZ theorem if the statistical dependency between X and Y are available at the decoder described as $R^{WZ}(D)$ is bigger than the case where, the correlation are exploited at both encoder and decoder for the same average distortion D. Therefore, the WZ theorem can be described as,

$$R^{WZ}(D) \geq R_{X|Y}(D), D \geq 0 \tag{1.8}$$

where, $R^{WZ}(D)$ is the minimum WZ encoding rate and $R_{X|Y}(D)$ is the minimum encoding rate of X, where Y is simultaneously available both at the encoder and decoder.

So when D=0 or no distortion exists the WZ theorem is similar to SW theorem. On the other hand, a reconstruction of information X is resulted that consists of an arbitrarily small error probability even if the source correlation between X and Y are exploited only at the decoder.

1.1.3 Some promising DVC applications

This sub-section presents some of the potential applications of DVC in order to understand the motivation and the driving force behind the DVC research.

(a) **Wireless low power surveillance**

It is the process of monitoring the behavior of people, objects or processes within the systems for conformity to the expected or desired norms in trusted systems for security. In the wireless low power surveillance the multiple cameras sensing the same event from different locations. The different encoders sense partially over-lapping areas and therefore their associated video sequences are correlated. In this case, the number of encoders is usually much more higher than the decoder (typically one). So it reduces the cost of the system. WZ coding in DVC has the advantage of resulting in low complexity encoders which helps in reducing complexity and power consumption. Wireless low power surveillance can be used to monitor the activity in private and public spaces. It also helps in military aspect to collect the information about the enemies. Figure 1.7 shows a realization of wireless low power camera surveillance.

Figure 1.7: Ordinary wireless camera (left) and wearable wireless webcam (rights) for surveillance

(b) **Wireless mobile communication**

Wireless mobile communication is one of the important application of DVC. In this case, both the devices have limited power and computational

resources. Figure 1.8 shows the wireless mobile communication between a pair of camera phone. In this scenario, power consumption and battery life are the key issues. So very low cost DVC encoder and equally low complex conventional decoder are placed. However, to take the advantage from DVC a high complexity decoder is placed at the base station. The DVC bit stream are sent to the base station, the base station or transcoder encodes the DVC bit stream either MPEG-X or a H.26X bit stream and transmits to another low complexity decoder.
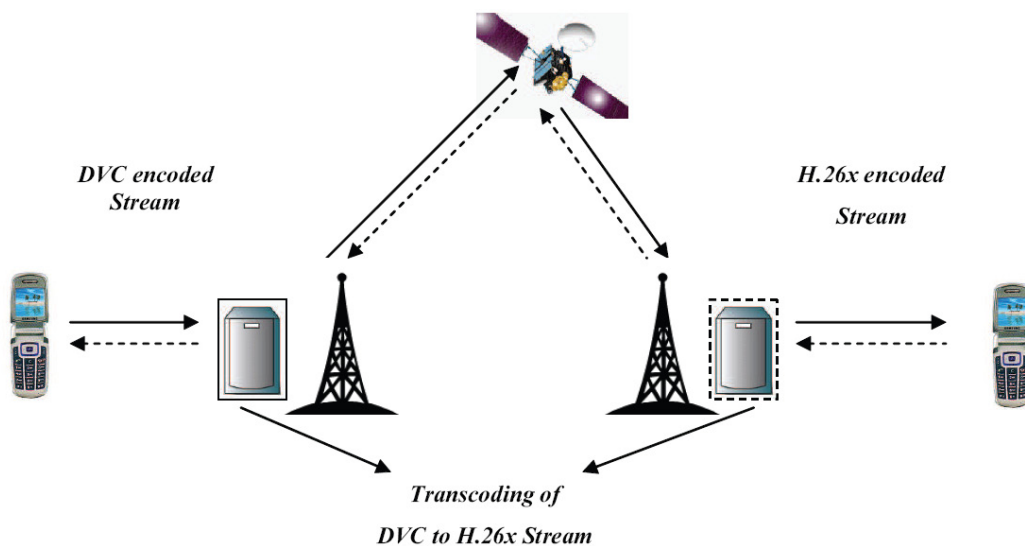
Figure 1.8: Wireless mobile communication scenario

(c) **Video sensor networks**

Another potential application of DVC is video sensor networks. In this case, multiple sensor nodes are available making some computer vision tasks (e.g. gesture recognition, wild life monitoring) etc. DVC is applicable for this video based sensor networks since it allows the construction of low complexity, low power encoder devices. The decoder is a high computational device capable of processing jointly to all information received by the encoder. Figure 1.9 shows an application scenario of video based sensor network.
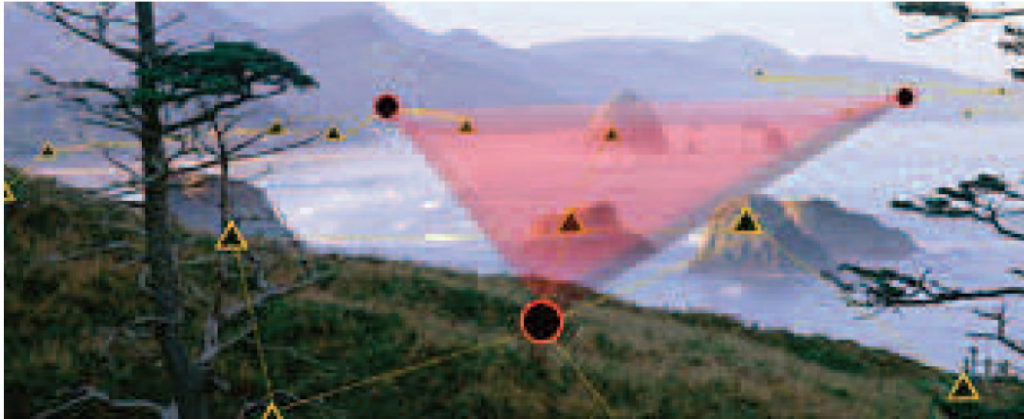
Figure 1.9: Wild life monitoring using video sensor networks

1.2   Review of distributed video coding

In recent years, DVC is becoming more popular due to the emerging applications like mobile camera phone, video surveillance, wireless camera. So it is a challenge for traditional video coding to fulfill the requirements of the above mentioned applications. So DVC is an ultimate choice over traditional video coding where the computational complexity has been shifted from encoder to decoder. First, we have reviewed some of the image and video coding approaches where DSC is applied successfully.

The DSC principles of SW theorem and WZ theorem were put forward in 2002 and becomes popular for last few years. In 1999, Pradhan and Ramchandran addressed the asymmetric case of source coding in which the binary and Gaussian sources are statistically dependent [11]. They have used scalar and trellis coset construction and SI is available at the decoder. Authors in [12–14] have considered the symmetric cases in which source and SI are encoded with the same rate. Wang and Orchard [12] made improvements over [11] by considering the asymmetric coding of Gaussian sources employing the embedded trellis code structure.

In 2002, Liveris et al. have applied turbo codes to encode the images that exhibits nearly Gaussian correlation between co-located pixel values [15]. In 2002, turbo code based compression scheme is applied for statistically dependent binary sources [13,14,16]. The turbo code based compression schemes can also be applied

to statistically dependent non binary symbols [17, 18], Gaussian sources [16, 19] as well as single sources [20, 21]. With the invention of iterative channel codes, it is applied for joint source channel coding case. Here, both the statistics of the source and channel are available at the decoder [16, 20–22]. In 2003, another powerful code known as low density parity check code (LDPC) is introduced. LDPC is a powerful alternative of turbo code [23–26]. Both turbo code and LDPC code are successfully applied to SW theorem and show improved performance results.

After the invention of some powerful channel codes like turbo code and LDPC code the practical work towards the DVC or WZ video coding becomes active. In 2003, Ahron et al. have proposed a low complexity WZ video coding named as distributed compression for large camera arrays. In this scheme, multiple correlated sequences are encoded with pixel domain WZ codec but are jointly decoded at central decoder [27]. Here, authors have made a comparison between WZ pixel domain architecture with JPEG 2000, a wavelet based coding method. In [28], the authors have proposed a transformed domain WZ coding solution rather than pixel domain, which leads to a better coding efficiency. In 2004, the authors have presented an intra-frame encoder and inter-frame decoder WZ video coding [29]. In this scheme, the encoder supplies some additional information about the current frame to help the decoder in the motion estimation task.

The application of DSC to DVC has evolved significantly over time in terms of coding framework improvement, improved compression efficiency. So several research community have evolved in this area. Various DVC architectures proposed by research groups are discussed below.

Recently, major practical solutions of DVC have been proposed by two groups: Bernd Girod's group at Stanford University [5] and Ramchandran's group at the University of California, Berkeley [30]. In 2002, the University of Stanford put forward a proposal for pixel domain DVC, introducing two different types of frames in a video called WZ frames and key frames [31]. This framework is commonly known as Stanford DVC framework. Later, they have proposed transform domain DVC framework [28], which results good compression efficiency as compared to pixel domain coding in DVC. Ramchandran's group at University of Barkeley have proposed a solution named as power efficient robust high compression syndrome

based multimedia (PRISM) coding. In this solution, it combines both the features of intra-frame coding with inter-frame coding compression efficiency [6, 30, 32, 33]. This architecture is quite different from Stanford based architecture in terms of SI generation and channel coding. Figure 1.10 shows the architectural description of PRISM codec.
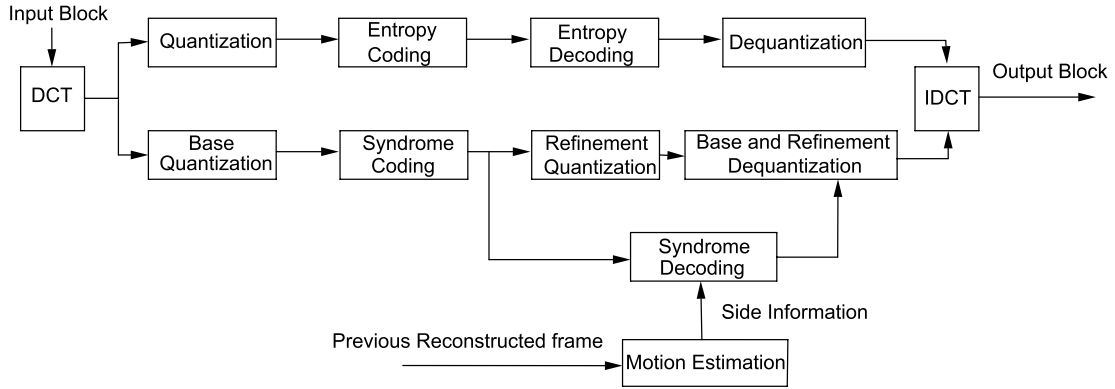
Figure 1.10: Architecture of PRISM

All our suggested schemes are based on Stanford based architecture and hence a detailed discussion is made for Stanford architecture.

The first practical solution towards DVC is pixel domain coding solution proposed by Girod's group at Stanford University [31, 34, 35] and shown in Figure 1.11. In this framework, the frames are divided into two groups, namely WZ frames composed of even numbered frames and key frames are the odd ones. If the temporal index of the video sequence is i ($i \geq 0$), the WZ frame X_{2i} is intra-frame coded at the transmitter and Y_{2i} is the inter-frame coded at the receiver with the help of SI. On the other hand, the key frames X_{2i-1} and X_{2i+1} are encoded with conventional intra-frame video coder. The SI, Y_{2i} at the decoder is generated by interpolating the two closest adjacent decoded key frames, one temporally in the past and another in future. The SI plays a key role in DVC architecture. It is treated as the noisy version of WZ frame X_{2i}. The detailed procedure of the pixel domain coding solution is given in following steps,

(i) The video frames are divided into WZ frames and key frames.

(ii) Each pixel in each WZ frame, X_{2i} is scanned row by row and is quantized using 2^M level uniform quantizer, where $M = 2, 4, 8.....$
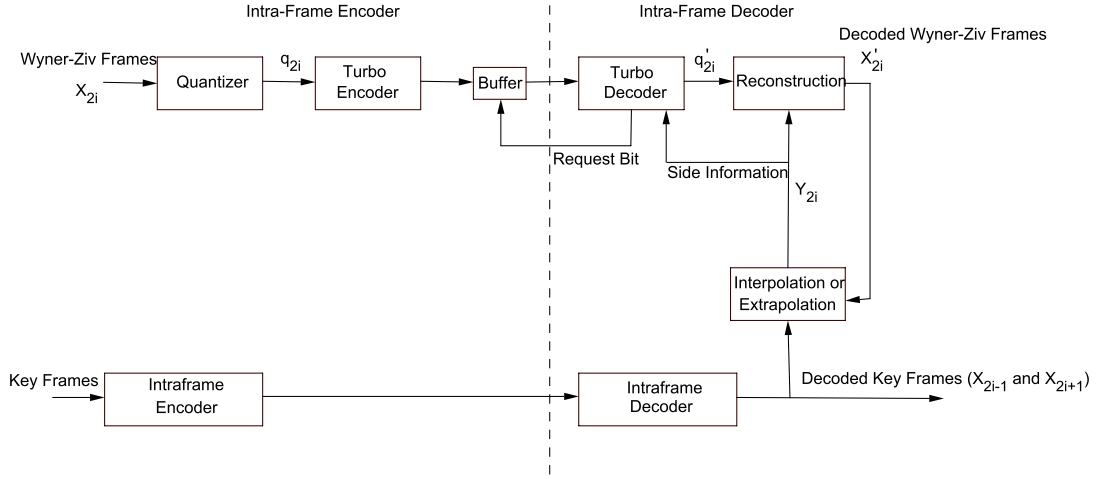
Figure 1.11: Stanford based pixel domain architecture

(iii) Turbo encoding is applied to the quantized symbol stream q_{2i} and the parity bits are saved into a buffer discarding all systematic bits.

(iv) At the other side of the DVC encode key frames X_{2i-1} and X_{2i+1} are transmitted via conventional intra-frame video coder.

(v) Decoder uses the frame interpolation technique to generate the estimated WZ frame known as SI and denoted as, Y_{2i} frame from the two adjacent key frames X'_{2i-1} and X'_{2i+1}.

(vi) A Laplacian distribution model is used to estimate the statistical properties between original WZ frame (X_{2i}) and SI (Y_{2i}) frame.

(vii) The received parity bits and the derived Laplacian distribution parameters are used by the turbo decoder to obtain the quantized symbol stream q'_{2i}.

(viii) Initially the decoder requests a fraction of parity bits from the encoder buffer to perform turbo decoding. The decoder also determines if the current bit error probability, P_e exceeds 10^{-3}, then decoder requests more number of parity bits. This process is repeated till successful decoding is completed.

(ix) The SI frame and quantized symbol stream q'_{2i} are used together to reconstruct the complete WZ frame X'_{2i}. Mathematically, this can be

expressed as

$$X'_{2i} = E(X_{2i} \mid q'_{2i}, Y_{2i}) \tag{1.9}$$

where, X'_{2i} is the reconstructed WZ frame, Y_{2i} is the SI, $E(\cdot)$ is the expectation operator, and X_{2i} is the original WZ frame.

The complexity of the pixel domain coding solution is simplest because neither discrete cosine transform (DCT) nor motion estimation is required. However, it requires a simple quantizer and a channel coder. Use of transform coding is an established concept in video coding. It helps to improve the compression efficiency by introducing the DCT with the expense of a little added complexity. DCT has been most popularly used in many video coding standards [28,36]. However, some researchers have introduced wavelet transform [37] in context to video coding. In transform domain Wyner-Ziv (TDWZ) video coding solution the DCT and bit plane extraction process are introduced. Figure 1.12 shows the overall architecture of Stanford based TDWZ video coding. The detail stepwise procedure of TDWZ video coding is described as,
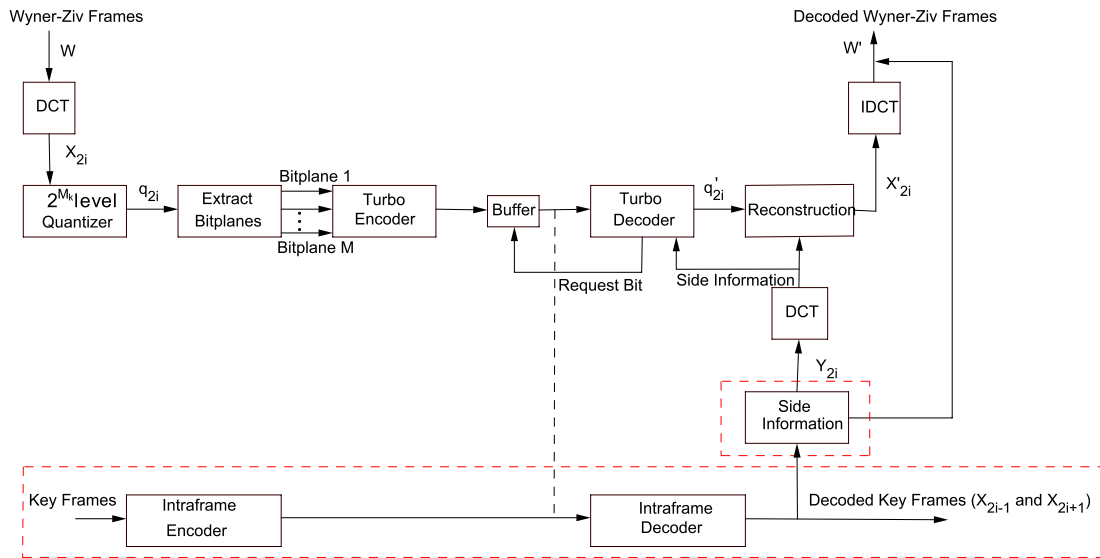
Figure 1.12: Stanford based transform domain architecture

  (i) The frames in a video sequence are divided into WZ frames and key frames.

  (ii) For each WZ frame X_{2i}, a 4×4 block wise DCT is applied.

(iii) The transform coefficients (DCT coefficients) of the frame X_{2i} are grouped together. The coefficients from the same position of each DCT blocks are used to compose 16 possible DCT coefficient bands.

(iv) Each DCT coefficient band X_k, $k = 1,2,...,16$ is uniformly quantized to obtain quantized symbol stream q_{2i}.

(v) The different quantized coefficients of the same band are grouped together and different bit planes are extracted. The bit planes are organized from the most significant bit (MSB) plane to least significant bit (LSB) plane.

(vi) Turbo encoding is applied to each bit plane. The turbo encoder generates the parity bits for each bit plane which is saved in a buffer and send to the decoder upon request. A pseudo random puncturing pattern is used to transmit the parity bits.

(vii) Meanwhile, the key frames are encoded using the conventional intra-frame video coding.

(viii) Decoder uses the frame interpolation technique for estimation of X_{2i} frame known as side information, Y_{2i} frame from the two adjacent key frames X_{2i-1} and X_{2i+1}.

(ix) The same block wise 4×4 DCT is applied to the interpolated frame, Y_{2i} to generate an estimate of X_{2i}. The correlation between corresponding coefficient band X_{2i} and Y_{2i} is modeled by a Laplacian distribution.

(x) The decoder also determines if the current bit plane error probability, P_e exceeds 10^{-3}, the decoder requests more number of parity bits; other wise, the current bit plane is executed successfully.

(xi) After all bit planes are executed successfully and the quantized symbol stream q'_{2i} is obtained, then reconstruction of each DCT coefficient band is resulted.

(xii) After all DCT coefficient bands are reconstructed, an inverse discrete cosine transform (IDCT) is applied to obtain the X'_{2i} frame.

The TDWZ video coding performs better performance as compared to PDWZ solution due spatial redundancy exploited by DCT. However, the quantization and reconstruction modules are different in TDWZ coding. The detail description of transformation, quantization, turbo encoding, feedback channel etc. are discussed in [3, 7].

Instituto Superior Tecnico (IST) group is one of the leading group in the field of DVC. Most of the solutions of IST group in the field DVC exhibit a strong potential for new applications targeting new advances in coding efficiency, error resilience and scalability. The IST group uses the same Stanford based architecture for pixel domain, referred to as Instituto Superior Tecnico pixel domain (IST-PDWZ) solution as well as for transform domain referred to as Instituto Superior Tecnico transform domain (IST-TDWZ) solution. Some of the major improvements of DVC by IST group are reviewed and compared with Stanford based architecture below.

(a) **IST pixel domain distributed video coding**

The first attempt towards IST-PDWZ solution is proposed by Ascenso et al. in 2005 [38]. The IST-PDWZ video coding is based on the same architecture proposed by Stanford group. However, there are two major improvements in terms of bit plane transmission and SI generation. Figure 1.13 shows the architecture of IST-PDWZ.
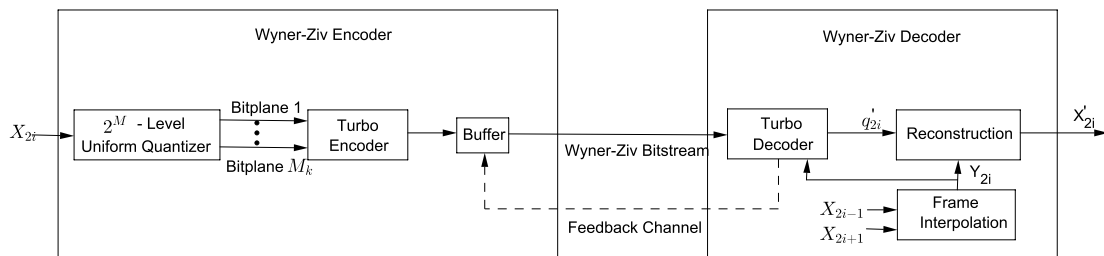
Figure 1.13: Architecture of IST-PDWZ

The bit plane extraction process is used by Stanford based TDWZ architecture, however, in PDWZ it utilizes the concept of bit plane extraction process. In this scheme, each pixel of WZ frame X_{2i} is scanned row wise and quantized using 2^M level of uniform quantization. After that, bit plane

extraction process is performed by grouping the bits of quantized symbols from the MSB plane to LSB plane. Then each bit plane is independently turbo encoded, starting with MSB to LSB. Authors in [39] have proposed an inverse order of bit plane transmission sending the LSB to MSB plane. In [28, 31], authors have proposed a PDWZ solution with or without bit plane transmission and claimed that the overall effect is similar. However, the bit plane extraction process is useful in DVC as it reduces the size of the input block for turbo encoder, thus reduces the decoding complexity.

The SI generation process in IST-PDWZ is completely different from Stanford based PDWZ. In this scheme, the IST researchers have proposed a sophisticated and complicated SI framework, where so many modules are associated to generate a good SI. Although it is complicated it shows improved performance. Figure 1.14 shows the SI generation framework in IST-PDWZ.
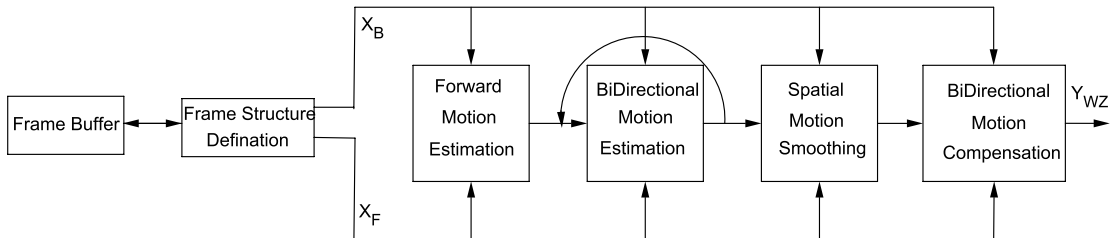
Figure 1.14: SI Generation module in IST-PDWZ

To remove noise, key frames are passed through lowpass filter. In this module, both the key frames are lowpass filtered and helps to remove noise and improve the accuracy of the motion estimation. A full search block matching algorithm is employed between decoded X_{2i-1} and X_{2i+1} key frames. The block based motion estimation algorithm is used due to its low complexity as compared to other algorithms. A bi-directional motion estimation is applied after a forward motion estimation to search two blocks that match most. This fixes the bi-directional motion vector to be used during interpolation. The resulted motion vector from bi-directional motion estimation suffers from low spatial coherence. So weighted vector median

filter is used to improve the quality of SI. Finally, a bi-directional motion compensation is performed.

(b) **IST transform domain distributed video coding**

The architecture of IST-TDWZ (Figure 1.15) is same as that of Stanford TDWZ. The SI generation process is also same as IST-PDWZ.
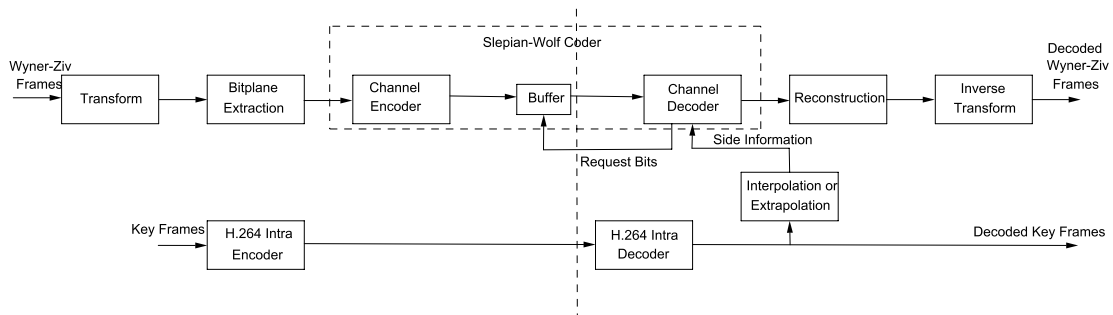
Figure 1.15: Architecture of IST-TDWZ

During the quantization of DCT coefficients, the value range of the AC coefficient bands is known to the decoder. With this operation, the quantization loss can be minimized. If the same number of quantization levels is applied to a value range which is smaller than a fixed value range, then a smaller quantization interval means that a smaller quantization step size is resulted. Therefore, a lower distortion of output can be obtained.

1.3 Other advances in distributed video coding

In DVC architecture different modules like transformation, quantization, channel coding, side information, intra-key-frame coding, feedback channel etc. are present which affect the overall performance of DVC. So this section reviews some of the notable advances on different modules in DVC.

(a) **Advances in use of transforms**

In most of the DVC framework discrete cosine transform (DCT) is frequently used and shows improved performance results. However, some researchers have extensively used the wavelet based transformation in DVC. There are several algorithms employed in wavelet based coding out of which embedded

zero tree wavelet (EZW) [40] and set partitioning in hierarchical trees (SPIHT) are most popular to encode the wavelet coefficients [41]. The first proposal towards discrete wavelet transform (DWT) is presented by Feng et al. in [42]. In this scheme, the spatial and signal noise to ratio (SNR) scalability are explored and they have claimed the proposed system is robust in channel error. Later, Magli et al. [43] and Wu et al. [44] have proposed the DWT based scalability method to further improve the performance. Authors in [45] have also addressed the scalability features in context to DWT. In 2006, Guo et al. have proposed a WZ video coding using wavelet where, EZW is used to encode the wavelet coefficients [37]. In this solution, the wavelet quantized coefficients are represented in terms of zero tree structure to determine the insignificant and significant coefficients. The significance map is intra coded and transmitted. The significant coefficients are independently encoded with turbo code and only the parity bits are transmitted. In [46], a WZ video coding scheme based on DWT and SPIHT algorithm is used to exploit the spatial, temporal, and statistical correlation of the frame sequence. In this work, DWT is used before quantization, then only low frequency components are WZ encoded using turbo code and the high frequency components are coded using SPIHT algorithm. This scheme performs better than intra-frame coding using SPIHT algorithm only. Tonmura et al. have proposed a DVC based JPEG 2000 coding scheme [47]. Here, the key frames are coded using JPEG 2000 intra coding. The WZ frames are coded using gray code and optimum quantization is used to improve the resolution and quality. This scheme achieves 5 dB PSNR gain as compared to the conventional JPEG 2000 coding. Wang et al. have proposed an efficient hybrid DVC [48]. In their work, they have adopted a wavelet based WZ codec to compress the residual frame between current and its reference frame. An intra mode decision based on temporal and spatial correlation is employed to determine whether a wavelet block should be coded using intra-SPIHT coding or SW-SPIHT coding. In 2006, Cheung et al. have proposed an efficient wavelet based predictive SW coding for hyperspectral image [49]. In this scheme, DWT is applied on WZ frame

and for channel coding they have used LDPC code instead of turbo code. They have applied the SPIHT algorithm to encode the wavelet coefficients.

In 2008, Ponchet et al. [50] have presented the second generation wavelets in DVC. In this framework, the WZ frames are encoded using turbo codes and DWT. This approach gives a great encoding runtime reduction.

(b) **Advances in quantization**

In WZ video coding solutions, quantizer plays a significant role to improve the coding efficiency. In most of the DVC solutions, uniform scalar quantization is applied. However, sophisticated quantizers are good options to further improve the DVC performance. The only problem to use some sophisticated quantizer in DVC is that, it increases the encoder complexity to a great extent. Wang et al. have proposed a lattice vector quantizer (LVQ) in WZ video coding [51]. For channel coding they have used LDPC code. This framework shows 1 dB PSNR improvement than scalar uniform quantizer. It also preserves the property of low complexity encoding. The similar work also can be found in [52].

In [53], the authors have presented a non-uniform scalar quantizer in DVC. In this algorithm, a probability distribution model is used which consider the influence of the joint distribution of input source and SI. Then, a modified Lloyd-Max algorithm is used to design a scalar quantizer to give an optimal quantization. Here, they have claimed an improved coding performance at low bit rate. Authors in [54], have extended the Lloyd-Max quantizer. Xu et al. have proposed a nested scalar quantizer in practical layer WZ video coding [55]. This scheme is a new technique for video streaming over wireless networks. In 2007, Shing et al. have proposed an adaptive nested scalar quantization scheme in DVC [56]. In this work, the absolute frame difference between current WZ frame and previous key frame are utilized, then an adaptive quantizer step size is decided according to threshold and sent to the decoder. This scheme exhibits higher rate distortion (RD) performance for low motion video. Weerakkody et al. in 2009 have presented a non-linear quantization for pixel domain DVC [57]. This scheme has better PSNR

improvement as compared to pixel domain DVC that uses linear quantizer. In 2010, authors have addressed a method for modeling, analyzing, and designing WZ quantizer for jointly Gaussian vector data with imperfect SI [58]. Zhang et al. have presented the quantizer design for correlation noise in DVC. In this work, they have designed a non-uniform quantizer using Lloyd-Max algorithm as well as dead zone scalar quantizer to enhanced RD performance in DVC [59]. In 2012, the authors have proposed a multimode nested quantizer in presence of uncertain SI and feedback [60]. In this scheme, the quantization parameter, feedback scheme, and source coding rate are jointly optimized to minimize the average rate or distortion.

(c) **Advances in Slepian-Wolf codec**

Turbo coding is the most common coding tool used in SW coder in Stanford architecture. Ramchandran's et al. in 2002 have addressed an alternative to turbo code known as LDPC code [61]. It has the capability to achieve the SW achievable boundaries. It is a new powerful channel code that has the ability to show similar performance as compared to turbo code. Several authors have reported and adopted LDPC code which works in the similar way as turbo code [28, 49, 51, 55, 62]. The quantized symbol stream followed by bit plane extraction process is sent to the LDPC encoder. The LDPC encoder generates the syndrome bits and send to the decoder. The decoder decodes the frame with the help of SI and syndrome bits. However, it is difficult to made any conclusion that LDPC code shows superior performance than turbo code for WZ video coding solutions. In 2006, Wang et al. have proposed a LVQ based DVC with LDPC code in pixel domain coding solution [51]. Westeriaken et al. have addressed a solution to analyze symbol and bit plane based LDPC in DVC and made a conclusion that both codes have similar performance [63]. More specifically it is fair to say that LDPC code is widely used in PRISM architecture, where, a turbo code is used in Stanford based architecture.

In 2009, the authors have addressed a non-uniform LDPC for multi biplane in DVC [64]. Recently, Li et al. have presented an improved LDPC

coding scheme with motion decision for DVC [65]. In this framework, the authors have claimed a PSNR improvement of 0.6 dB gain as compared to conventional LDPC using DISCOVER. Brites et al. in 2012 have proposed an augmented LDPC graph for DVC with multiple SI [66]. In this work, multiple SI hypothesis are available at the decoder. The advantage of this scheme is to exploit multiple SI through an efficient joint decoding technique with multiple LDPC syndrome decoder. It exchanges the information to obtain the coding efficiency to a greater extent.

(d) **Advances in correlation noise modeling**

The correlation noise modeling is another important module in DVC. Without appropriate noise modeling the decoding output of DVC will give a wrong result. In most of the DVC solutions a Laplacian distribution is used to calculate the residual statistics between WZ frame and SI frame. So the parameters are usually estimated offline. In most of the solutions the Laplacian parameter can be estimated using the SI and one of the adjacent key frames or one previously decoded WZ frames [5, 28, 67]. In Brites et al. have proposed a correlation noise modelling for efficient pixel and transform domain WZ video coding [68–70]. In this scheme, the estimation is done in 3 levels of granularity such as frame level, block level, and pixel level. In PDWZ, all this 3 levels are used where, in TDWZ coding, 2 levels granularity are used. The higher the estimation granularity the better is the RD performance. In 2006, Dalai et al. have presented a report on improving the codec integration through better correlation noise modeling [71]. In this scheme, an improved model is used for the correlation between SI and original WZ frame. In this work, they have made a conclusion that by modeling the non-stationary nature of the noise leads to substantial gain in the RD performance. In [72], authors have proposed the SI quality varies from point to point due to the content of the video frame nature (e.g occlusion). The same authors have presented a related discussion on dynamic estimation of the virtual channel model [73]. Guo et al. have addressed a new pixel domain DVC to exploit both temporal and spatial correlations

at the decoder [74]. In this scheme, a new probability model is proposed in which transitional probability is calculated from the conditional probabilities on the multiple SI signals. However, authors in [75], have reported a wavelet based DVC to model the correlation between wavelet sub-bands. Later, in 2009, Li et al. have presented a multi-view DVC to model the correlation statistics at decoder [76]. In 2010, Tsai et al. have proposed a practical estimation of adaptive correlation noise model for DVC [77]. In this work, the authors have suggested that, in order to retain the low cost and an efficient system a SI refinement with practical correlation noise modeling estimation is required.

(e) **Advances in reconstruction**

The basic reconstruction model was presented in the WZ video coding solution by Girod et al. in [31]. After that, some improvement have been incorporated for this purpose. More enhanced approaches for this purpose have been presented in [78, 79] known as minimum mean square error (MMSE) techniques. In 2007, Vatis et al. have reported an average distribution of the transformation coefficients and obtained a coding gain of 0.9 dB [78]. Kubasov et al. in [79] have proposed the problem of optimal mean square error (MSE) reconstruction of quantized samples in WZ video coding system. In this scheme, they have developed an algorithm to compute the reconstructed value x' as,

$$x' = E\left[x \mid x \in \left[(z^-, z^*), y\right]\right] \qquad (1.10)$$

where, x is the random variable, z^-, z^* are the reconstruction boundaries and E is the expectation.

In 2009, Liu et al. have addressed a novel two-pass reconstruction algorithm for DVC, in which the traditional reconstructed WZ frame is utilized to perform motion estimation to obtain a more accurate motion field [80]. Authors in [81], have presented a novel reconstruction algorithm in which pixel values are thought as discrete random variables. The experimental results shows the superiority as compared to optimal reconstruction

algorithm in terms of RD performance.

In 2010, Badem et al. have reported a novel transform domain uni-directional DVC (UDVC) without feedback channel [82]. This scheme, works well for low complexity DVC codec in order to eliminate the feedback channel of the existing DVC and simple encoder rate control algorithm is used.

(f) **Advances in side information**

Side information is most widely focused research area in DVC. It is one of the key information that directly influence the DVC performance. Initially, Girod et al. have proposed two hierarchical frame dependency arrangement [35]. In their first approach the SI for the current WZ frame can be extrapolated from a key frame or from a WZ frame. In their second approach, a more complex arrangement has been used with an increase temporal resolution of 2:1 with bi-directional interpolation.

Later in [28], they have proposed motion compensated interpolation (MC-I) and motion compensated extrapolation (MC-E). In MC-I, the SI for an even frame at time index t is generated by performing motion compensated interpolation using the decoded two closest key frames at time $(t - 1)$ and $(t + 1)$. In MC-E, the SI can be generated by estimating the motion between the decoded WZ frame at time $(t - 2)$ and decoded key frames at time $(t - 1)$.

In [28] Girod et al. have also proposed previous extrapolation (Prev-E) and average interpolation (AVI) technique to generate SI for low complexity video coding solution. In Prev-E scheme the previous key frame is used directly as SI. In AVI technique the SI for the WZ frame is generated by averaging the pixel values from the key frames at time $(t - 1)$ and $(t + 1)$. Natario et al. [83] have proposed a motion field smoothing algorithm to generate the SI. This solution is designed for pixel domain coding architecture.

Artigas and Torres have proposed an iterative motion compensated interpolation technique where the turbo decoder runs several times for decoding the WZ frame to generate an estimated side information [84].

Adikari et al. have proposed a multiple SI stream for distributed video coding [85]. It uses two SI streams which are generated using motion extrapolation and compensation ME-C. The first SI stream (SS-1) is predicted by extrapolating the motion from the previous two closest key frames. The second SI stream (SS-2) is predicted using the immediate key frame and closest WZ frame.

In [86] Fernando et al. have proposed a SI scheme using sequential motion compensation, using both luminance and chrominance information to improve the decoding performance of DVC. The extension of this work have been made by Weerakkody et al. in [87]. In this scheme a spatio-temporal refinement algorithm is used to improve the SI resulting from motion extrapolation.

In [88] Badem et al. proposed a novel SI refinement technique based on motion estimation in the DC domain for transformed domain based DVC. Varodayan et al. [89] have proposed an unsupervised motion vector learning. This method applied an expectation maximization (EM) algorithm for unsupervised learning of motion vectors. The authors have claimed a better RD performance.

Brites et al. proposed a frame interpolation framework with forward and bi-directional motion estimation with spatial smoothing [38]. The most promising SI refinement framework has been adopted and extended in [90, 91]. In this framework, first both the key frames are low-pass filtered. Then a block matching algorithm is used to make the motion estimation between two adjacent key frames. The bi-directional motion estimation is performed followed by the forward motion estimation. After the bi-directional motion estimation, a spatial motion smoothing algorithm is employed. Once the final motion vectors are obtained, the bi-directional motion compensation is performed.

(g) **Advances in intra-key-frame coding**

The coding performance of the WZ video coding strongly depends on the quality of the SI. The SI in DVC can be generated from the decoded key frames resulting from the intra-frame DVC coder. So it is necessary to have a good decoded key frames at the decoder side of the DVC. Very few works have been reported in the literature in intra-key-frame coding.

Girod et al. in [34] have proposed two hierarchical frame dependency arrangement, where the frames are encoded as intra (I) frame with a fixed quantization parameter using H.263 (Intra) coding. The scheme has a limitation i.e. motion compensation interpolation is less accurate and thus degrades SI quality.

In [83] Brites et al. have proposed an improved transform domain WZ video coding architecture. Here, the key frames are sent directly to the decoder without any compression and a higher bit rate has been achieved. In [85,86], the authors have claimed high RD performance, but a poor SI is generated at the decoder end due to lossless key frames without any compression. Recently a reported scheme which uses H.264 intra (main profile) is popular among researchers [90]. This scheme shows improved performance results. In this suggested scheme, the key frames are always encoded with H.264 intra (main profile) and the quantization parameters (QP) for each RD point are chosen. The improvement in intra-key-frame coding in DVC is an unexplored area. There exists scope for improvement of intra-frame DVC coding for key frames.

(h) **Advances in DVC based video communication**

In most of the cases the transmission channel noise in DVC is ignored totally and most of the aforementioned schemes are assumed to be noiseless. However, Pedro et al. [92] for the first time have proposed a packet based network where it intends to study the error resilience performance of a feedback channel based transform domain DVC. Several aspects of video compression and transmission in wireless broadband networks are discussed

in PRISM architecture [93].

1.4   Motivation

It is observed from the literature that DVC is a prominent area of research due to its vast applications in handheld devices with less memory and computing power. In addition, DVC performance is associated with mostly on side information generation, which in turn dependent on reconstructed key frames. From the through investigation from the reported literature it has been observed that there exists a scope for improvement of DVC performance through better quality side information. Hence, we are motivated to propose few schemes for qualitative key frame generation and side information generation. The objectives are narrowed to –

(i) propose a Burrows-Wheeler transform based H.264/AVC intra-frame (BWT-H.264/AVC (Intra)) coding. The intra-frame coding in DVC is used to generate the side information. Hence the scheme thrusts to generate improved decoded key frames which in turn helps to generate improved side information.

(ii) develop a dictionary based intra-key-frame coding in DVC. The dictionary based approach replaces the conventional DCT based intra-frame coding. The objective is to generate superior quality decoded key frames.

(iii) devise an improved SI framework using multilayer perceptron (MLP-SI) scheme in DVC to improve the overall coding performance and visual perceptual quality.

(iv) propose MSVR based side information generation (MSVR-SI) scheme for overall improvement in DVC performance.

1.5   Thesis Layout

The thesis is organized as follows —

Chapter 2 : BWT based H.264/AVC intra-frame video coding (BWT-H.264/AVC (Intra))

The performance of DVC strongly depends on the quality of SI. The SI generation process estimates the WZ frame from the decoded preceding and succeeding key frames. Hence, high quality of decoded key frames results to superior SI generation. In this chapter, an investigation has been made to propose a Burrows-Wheeler Transform (BWT) based intra-frame coding to generate improved decoded key frames. The suggested scheme is embedded in H.264/AVC (Intra) coding framework and used to code the key frames. Comparative analysis with other standard techniques in DVC reveals that the proposed scheme has better standing to its counterparts in terms of both coding efficiency and improved perceptual quality.

Chapter 3 : Dictionary based H.264/AVC intra-frame video coding using OMP (DBOMP-H.264/AVC (Intra))

This chapter presents a dictionary based intra-frame video coding technique with adaptive construction of over complete dictionary. The traditional transform based intra-frame video coding is replaced by a dictionary based approach. The dictionary is completed using K-singular value decomposition (K-SVD) based offline training using residual intra coded macroblock of size 4×4 selected from different video sequences with different motion characteristics. For encoding the dictionary elements orthogonal matching pursuit (OMP) algorithm has employed. Finally, the overall performance of DVC is compared with other competent schemes.

Chapter 4 : An improved side information generation using MLP (MLP-SI)

This chapter presents an improved SI generation in a DVC framework. The scheme utilizes a multilayer perceptron (MLP) to predict the WZ frames. It accepts 8×8

blocks from predecessor and successor key frames of a WZ frame and estimates the corresponding 8×8 block of WZ frame. The MLP is trained offline using training patterns selected from frames with numerous motion patterns. Further, these frames are considered from different video sequences. Subsequently, the trained MLP is used for SI generation. The proposed scheme is validated with respect to training convergence, rate distortion (RD) performance, peak signal to noise ratio (PSNR) performance, number of requests per SI frame, decoding time requirement etc. Comparative analysis have been performed with standard video codecs using standard video sequences. The proposed MLP-SI scheme has superior performance over the existing schemes with respect to both qualitative and quantitative measures.

Chapter 5 : MSVR based side information generation with adaptive parameter optimization (MSVR-SI)

This chapter proposes an improved side information (SI) generation scheme using multivariable support vector regression (MSVR). MSVR is formulated to suit our problem. Furthermore, the parameters of the MSVR are optimized through particle swarm optimization (PSO) technique using mean square error (MSE) as the fitness function. For generating the SI in DVC, MSVR model takes the non-overlapping 8×8 blocks of two decoded key frames as inputs and predicts the 8×8 block of Wyner-Ziv (WZ) frame. The training is made offline and prediction is made online. The proposed scheme shows superior performance in terms of both coding performance and improved perceptual quality as compared to its competent schemes and conventional MSVR model.

Chapter 6 : Hybrid schemes formulation out of the suggested schemes

This chapter deals with the hybrid schemes generated out of our suggested schemes. Out of two intra-key-frame coding schemes (BWT-H.264/AVC (Intra), DBOMP-H.264/AVC (Intra)) and two side information generation schemes (MLP-SI, MSVR-SI), we combine the proposed intra-frame coding schemes with the side information generation schemes to create four different hybrid schemes in

permutation. The hybrid schemes are named as follows.

(a) **BWT-MLP** scheme — The scheme is formulated by combining BWT-H.264/AVC (Intra) intra-frame coding scheme and MLP-SI side information generation scheme.

(b) **BWT-MSVR** scheme — This scheme is a combination of BWT-H.264/AVC (Intra) intra-frame coding scheme and MSVR-SI side information generation scheme.

(c) **DBOMP-MLP** scheme — The scheme is an outcome of putting DBOMP-H.264/AVC (Intra) intra-frame coding followed by MLP-SI side information generation.

(d) **DBOMP-MSVR** scheme — This scheme deals with DBOMP-H.264/AVC (Intra) intra-frame coding and MSVR-SI side information generation together.

The hybrid schemes are incorporated into the Stanford based architecture of DVC. The performance analysis with respect to RD performance, SI-PSNR, overall PSNR, number of requests per SI frame, decoding time has been made to derive an overall conclusion.

Chapter 7 : Conclusions and future work This chapter provides the concluding remarks with more emphasis on achievements and limitations of the proposed schemes. The scopes for further research are outlined at the end.

Chapter 2

BWT based H.264/AVC intra-frame video coding

The performance of distributed video coding (DVC) strongly depends on the quality of side information (SI). The SI generation process estimates the WZ frame from the decoded preceding and succeeding key frames. The quality of SI depends on the accuracy of the decoded key frames. Hence, high quality decoded key frames results to superior SI generation. Coding key frames in DVC is largely an open issue, which has not been exploited to a great extent. The most popular intra-frame coding schemes such as JPEG-2000, MPEG-2, H.263 (Intra), and H.264/AVC (Intra) are used to transmit the key frames in DVC. However, this schemes fail to produce a good coding efficiency and visual quality results. In this chapter, an investigation has been made to propose a Burrows-Wheeler Transform (BWT) based intra-frame coding to generate improved decoded key frames. The suggested scheme is embedded in H.264/AVC (Intra) coding framework and used to code the key frames. Hence, the scheme is named as BWT-H.264/AVC (Intra). Comparative analysis with other standard techniques in DVC reveals that the proposed scheme has a better standing to its counterparts in terms of both coding efficiency and improved perceptual quality.

The rest of the chapter is organized as follows. Section 2.1 elaborates the related research on intra-key-frame coding in DVC. The theoretical foundations of BWT is discussed in Section 2.2. The proposed BWT-H.264/AVC (Intra) is presented in Section 2.3. Simulation results and comparative analysis of the suggested scheme are discussed in Section 2.4. Finally, Section 2.5 summarizes

the chapter.

2.1 Related research on intra-frame coding in DVC

This section presents the related research on intra-frame coding in DVC along with limitation of each scheme. Girod et al. have proposed two hierarchical frame dependency arrangement [28]. In this framework, the frames are encoded as Intra (I) frame with a fixed quantization parameter using H.263 (Intra) coding. This scheme utilizes motion compensation interpolation which is less accurate and thus degrades SI quality. Brites et al. have proposed an improved transform domain WZ coding architecture [83], where, the key frames are sent directly to the decoder without any compression and a higher bit rate is achieved. In [85,86], the authors have claimed high RD performance, but a high amount of bits are resulted from encoder to decoder due to lossless key frames without any compression. All the reported works have claimed about the improvement of DVC codec design, but none of the schemes address towards the improvement of intra-key-frame coding in DVC. For the first time, Adikari et al. have proposed an independent key frame coding using correlated pixels in DVC [94]. In this framework, a novel intra coding technique is proposed that eliminates the requirement of a secondary coding scheme for coding the key frames in DVC. The authors have claimed a peak signal to noise ratio (PSNR) improvement, but this scheme works in a pixel domain coding solution. Recently, most schemes in DVC deal with transform domain coding solution.

The latest framework adopted recently in [90, 91] are popular among DVC researchers, however, they do not address the improvement in intra-frame coding. In these schemes, the key frames are coded with the quantization parameters (QP) as defined in [91]. The QP are determined through an iterative process which stops when the average quality (PSNR) of the WZ frames is equal to the quality of the key frames.

So from the literature it has been observed that key frames are used as reference information in the process of SI estimation for the WZ frames at the decoder. So

there exists a need to introduce a good intra-key-frame DVC coder to produce a good compression efficiency at the encoder side and better visual decoded quality frames at the decoder to estimate the accurate SI. In this chapter, we propose a BWT based intra-frame coder integrated in H.264/AVC framework to improve the quality of decoded key frames.

2.2 Theoretical foundation of BWT

Transforming the original data to a transform domain sometimes make our problem simpler and lead to efficient solutions. A powerful transformation suggested by Michel Burrows and David Wheeler is widely used for data compression. The transformation is popularly known as Burrows-Wheeler Transform (BWT) and is reversible in nature, i.e. the transformed data can be restored without loss of fidelity [95].

The set of steps applied to an input stream in BWT are as follows,

1) write the input stream as the first row of a matrix one symbol per column.

2) form all cyclic permutations of that row and create the other rows of the matrix.

3) sort each row in lexicographic order.

4) consider final column as output of the sorted matrix along with the number of rows which correspond to the original input.

Move-To-Front (MTF) encoding is usually regarded as an essential part of BWT. MTF encoding assumes that symbols can be ranked according to their recency or the closeness of their last occurrence [96, 97]. Frequency of symbol is not an important issue but context of symbol is important. MTF encoding makes the symbols more regular to achieve better compression. The MTF keeps a list of the symbols in the source alphabet. For each new symbol, send the index of the symbol in the list and then move the symbol to the front of the list. Since the index of the first symbol in the list is zero, long runs of the same symbol will give

long runs of zero in the MTF output which could exploit the context adaptive variable length coding (CAVLC) to achieve higher compression.

BWT based compression method has been widely used and shows several improvements in text compression. Several authors have reported the improvements to the original BWT algorithm. In [98], Nilson et al. have proposed a scheme, where radix sort can be used as a first sorting step during the BWT. Fenwick has presented a block sorting text compression along with BWT which sorts long words instead of single bytes and shows compression improvement over conventional BWT [99].

On the other hand, very few selected contributions have been made on BWT in image compression. Unlike text compression BWT can not be used directly in image compression. After re-arrangement of pixels it can be exploited more efficiently. Guo et al. have proposed a waveform and image compression using BWT [100]. Annadurai et al. have addressed a color image compression using bit reduction and BWT [101]. They select blocks of pixel values from source image and convert them into equivalent alphabets or text. Then BWT based compression method can be applied to the text file to compress it. In [102], a lossy image compression based BWT in JPEG framework is used. In 2008, a lossless image compression using BWT is applied for medical image compression [103]. However, no attempt has been made to use the BWT in intra-frame video coding.

2.3 Proposed BWT based H.264/AVC (Intra) intra-frame coding

Among the variety of video coding schemes available today, the selection of an appropriate coding algorithm for a particular multimedia service becomes a crucial issue. In fact, the block based video coders seems to be more popular in multimedia services available today [104, 105]. Two different types of coding exist in a block based video coder, namely intra and inter coding modes. Intra mode treats a video as still image without any temporal prediction employed. In intra-frame coding mode, all macroblock of a frame are intra coded.

Normally, in a video frame the adjacent pixels are strongly correlated. So

the BWT utilizes this property and forms a good regularity structure with the help of MTF. However, BWT can not be applied directly to the video frames. After performing a reorder arrangement in the intra-frame coding it will provide a good regularity structure to achieve higher compression. Subsequently, a spatial directional prediction i.e. 9-mode prediction is applied to intra frames to improve the frame quality. In this scheme, BWT is embedded in H.264/AVC intra-frame coding and the overall steps followed are given in *Algorithm* 1.

Algorithm 1 : BWT-H.264/AVC (Intra) scheme

Step 1 : Perform the 9-mode intra prediction to predict the pixels in a macroblock.

Step 2 : Subtract the predicted macroblock from the current macroblock to generate the residual macroblock.

Step 3 : Apply a forward 4×4 integer transform to the residual macroblock.

Step 4 : Quantize the transformed coefficients and reorder them into an 1-D array using zig-zag scan to create the input stream for BWT.

Step 5 : Apply BWT to the stream followed by MTF to the last column of the generated matrix.

Step 6 : Pass the output of MTF to context adaptive variable length coding (CAVLC).

Step 7 : Apply the reverse MTF followed by reverse BWT to generate the original coefficients at the decoder.

Step 8 : Apply inverse transform and reconstruct the residual macroblock.
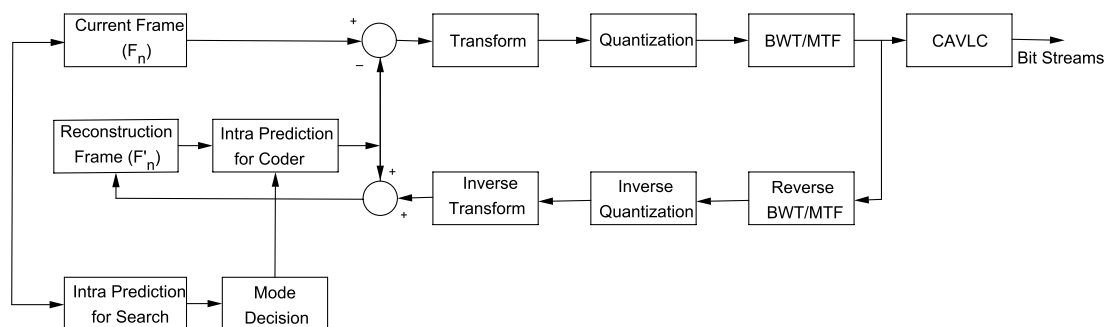
Figure 2.1: Block diagram of proposed BWT-H.264/AVC (Intra) scheme

The detail block diagram of the proposed approach is shown in Figure 2.1.

For completeness, let us have a detail description of proposed BWT-H.264/AVC (Intra) scheme. The intra prediction algorithm predicts and generates the pixel value in a 4×4 block from adjacent pixel values that have already been encoded. The H.264/AVC intra-frame encoder provides 9-mode prediction for 4×4 block. Generally, 4×4 prediction modes are used for highly textured region and 16×16 prediction modes for flat regions. The intra-frame predicted macroblock is subtracted from original macroblock to generate the residual macroblock. To reduce the correlation between the coefficients of a macroblock, the pixels are transformed into a different domain space by means of an integer transform. The quantizer is regarded as the most important component of the video encoder since it controls both the coding efficiency and quality of the reconstructed video sequence. In the proposed approach, the transform coefficients are quantized. Then, a zig-zag coefficient scanning method is used to convert the 2-D quantized matrix into 1-D matrix. The resulting output of zig-zag scanning will be put sequentially in a 1-D stream with the number of successive symbols. The output of the zig-zag scanning order is passed to BWT and a cyclic shift for the given data is made. The resulted cyclic permuted data is sorted in a lexicographic manner and the final column will be passed to MTF encoding. So for long runs of the same symbol will give long runs of zeros in the MTF output. The output of MTF is then passed to context adaptive variable length coding (CAVLC) to achieve higher compression efficiency. At the decoder side, the reverse CAVLC is performed followed by reverse BWT and reverse MTF. The quantized transform coefficients are used to generate residual information through inverse quantization followed by inverse transformation.

2.4 Results and discussions

In order to evaluate the efficacy of the proposed scheme, different experiments have been carried out on standard video sequences namely *Foreman*, *Miss America*, *Carphone*, *Coastguard* and *Silent*. Table 2.1 lists the main characteristics of each video sequence used. Among all, *Foreman* is the most complex sequence in terms of motion content. The proposed scheme has been tested in transformed domain

Table 2.1: Characteristics of video test sequences

Video Sequence	*Foreman*	*Miss America*	*Carphone*	*Coastguard*	*Silent*
Total Number of frames	400	150	400	400	400
Number of Frames Evaluated	200	150	150	250	250
Resolution	176×144	176×144	176×144	176×144	176×144
Temporal Resolution	15, 30	15	15, 30	15, 30	15, 30

WZ video coding architecture. For bitrate and PSNR performance evaluation, only luminance data is considered. Key frames as well as WZ frames are included in the rate distortion performance. No errors are introduced during transmission assuming WZ bit stream is error free. Eight quantization matrices are used as proposed in [91]. During the experiment the turbo encoder rate is assumed to be half. A Laplacian distribution model is used to exploit the correlation between original WZ frame and corresponding SI frame. To evaluate the overall performance of the proposed scheme the following codecs are used as the benchmarks.

(a) **H.263/Intra [28]** - H.263 Intra coding standard is not the best intra coding standard, but it includes most of the WZ video coding literature.

(b) **H.264/AVC Intra [105]** - The most popular intra coding codec till date that does not exploit the temporal redundancy.

(c) **H.264/AVC no motion (main profile) [90]** - Coding with H.264/AVC no motion (main profile) exploits the temporal redundancy but does not use any motion estimation.

(d) **IST-TDWZ [91]** - The most promising WZ video coding standard, which is currently considered as the state-of-art scheme with the best RD results.

The overall experimental study is divided into five different experiments and are discussed in sequel.

Experiment 1: Analysis of bits requirement per frame

The proposed scheme along with the traditional H.264/AVC (Intra) in IST-TDWZ scheme is simulated for *Foreman* and *Coastguard* video sequences using 150 frames from each video. The number of bits required per frame in both the schemes are computed separately and comparative performance is shown in Figures 2.2 and 2.3 for *Foreman* and *Coastguard* sequences respectively. It may be observed that there is a reduction in bits per frame requirement in both the video sequences, however, it is more significant in *Foreman* video. In particular, a reduction of 14.43 kbits and 7.55 kbits for few frames in *Foreman* and *Coastguard* are achieved. This shows the superiority of the proposed scheme as compared to H.264/AVC (Intra) used in IST-TDWZ approach.
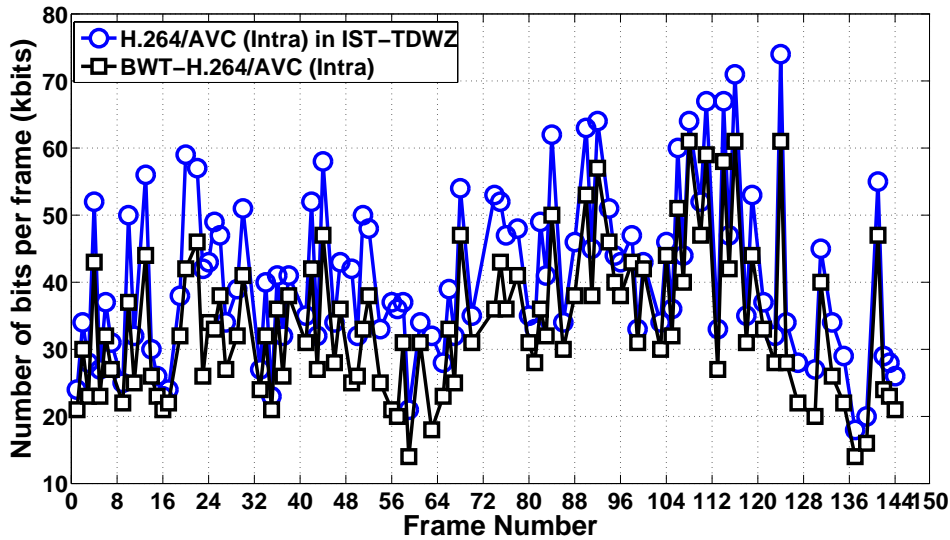
Figure 2.2: Number of bits requirement per frame (*Foreman*)

The subjective performance of the decoded key frames on different schemes are shown in Figures 2.4 - 2.6 for *Foreman*, *Coastguard*, and *Miss America* video sequences respectively. It may be observed that the proposed schemes works well for video with dynamic motion characteristics. The other schemes like H.263 (Intra) and H.264/AVC (Intra) suffers from blocking artifacts. Figure 2.4 shows the subjective analysis of 112th decoded key frames for *Foreman* sequence. The PSNR in (dB) of the decoded key frame for H.263 (Intra), H.264/AVC (Intra), and BWT-H.264/AVC (Intra) schemes are 28.55, 32.06, and 32.97 for *Foreman*
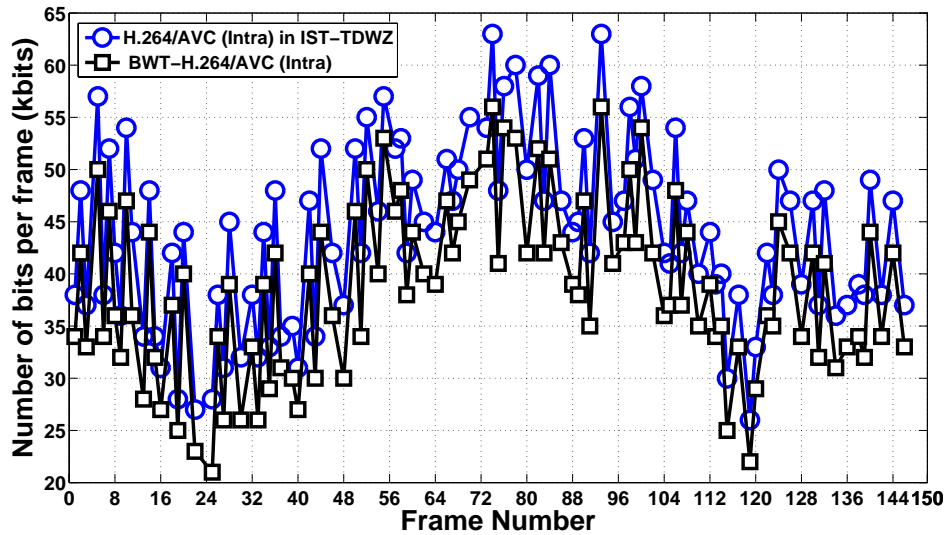
Figure 2.3: Number of bits requirement per frame (*Coastguard*)

(a) H.263 (Intra)

(b) H.264/AVC (Intra)

(c) BWT-H.264/AVC (Intra)

Figure 2.4: Subjective analysis of a decoded key-frame (112th frame of *Foreman*)

sequence. Similarly for *Coastguard* sequence the PSNR of decoded key frame are 26.44 dB, 30.37 dB and 32.54 dB for H.263 (Intra), H.264/AVC (Intra),

(a) H.263 (Intra)   (b) H.264/AVC (Intra)
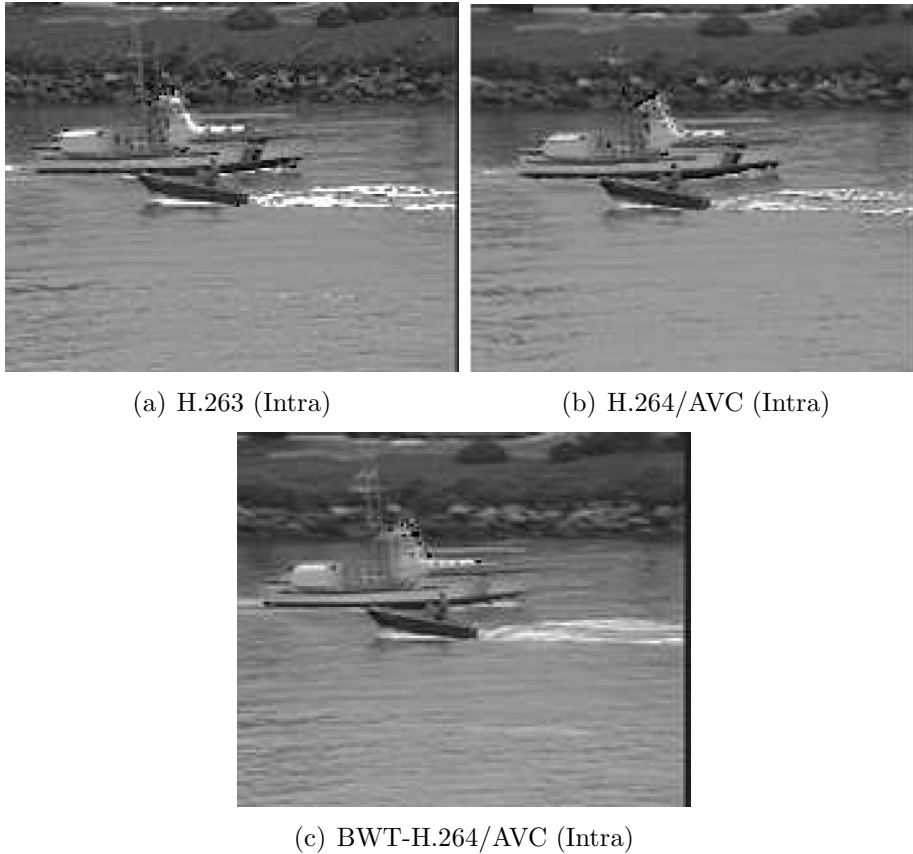
(c) BWT-H.264/AVC (Intra)

Figure 2.5: Subjective analysis of a decoded key-frame (35th frame of *Coastguard*)

BWT-H.264/AVC (Intra) schemes respectively. Further, the PSNR in (dB) for H.263 (Intra), H.264 (Intra), and BWT-H.264/AVC (Intra) are 28.2, 31.43, and 32.76 for *Miss America* sequence. So the proposed BWT-H.264/AVC (Intra) coder shows an improved decoded quality as compared to the best existing intra-frame coding schemes available in DVC.

Experiment 2: Analysis of rate distortion performance of key frames
The RD plots for different video sequences are shown in Figures 2.7 - 2.10 exhibit the comparison between proposed and traditional approach. Following inferences have been drawn from the observations.

(a) At 15 fps, the proposed BWT based H.264/AVC intra frame coder has better RD performance for *Foreman* and *Coastguard* sequences. In Figure 2.7 especially for *Foreman* sequence, the proposed scheme has an average PSNR gain of 0.42 dB as compared to conventional H.264/AVC (Intra) codec used in IST-TDWZ. For *Coastguard* sequence shown in Figure 2.8 the average

(a) H.263 (Intra)

(b) H.264/AVC (Intra)

(c) BWT-H.264/AVC (Intra)
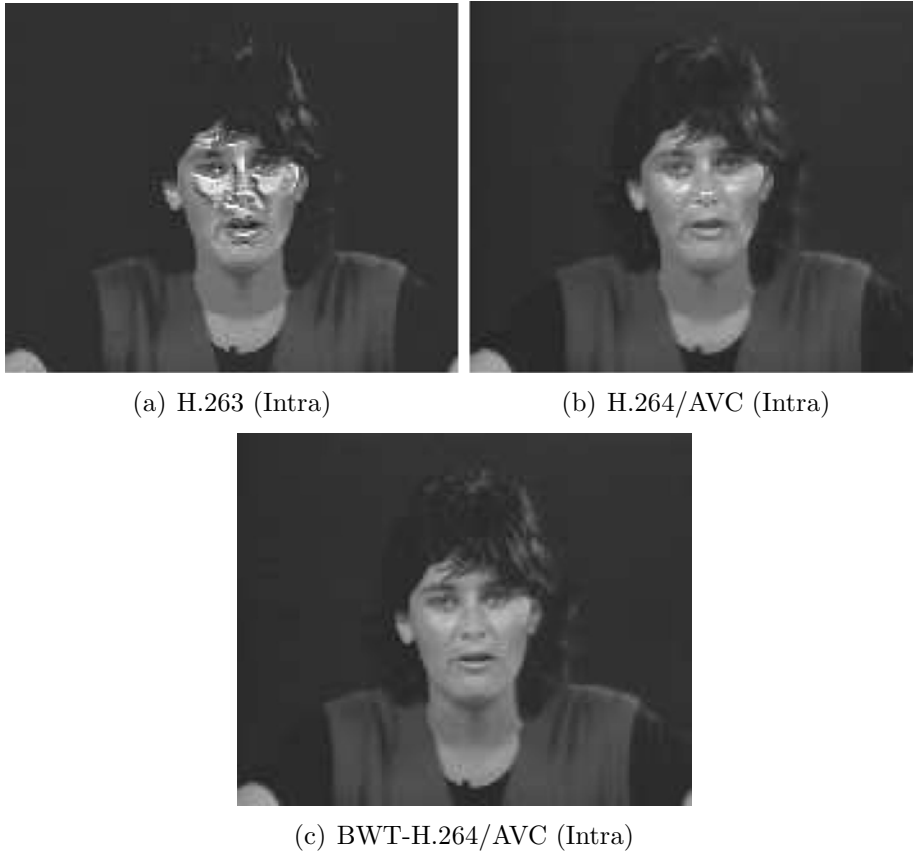
Figure 2.6: Subjective analysis of a decoded key-frame (84th frame of *Miss America*)

PSNR gain is observed to be 0.2 dB.
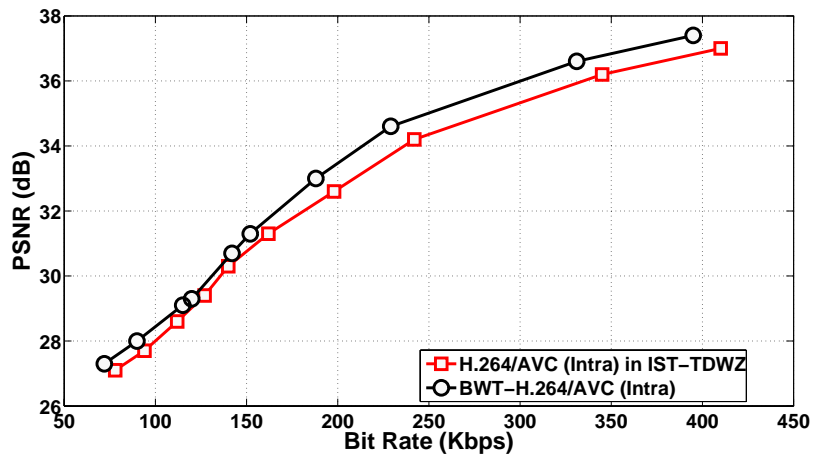
Figure 2.7: Rate distortion performance of key-frames (*Foreman* at 15 fps)

(b) Figures 2.9 and 2.10 depict the RD gain for sequences at 30 fps. The
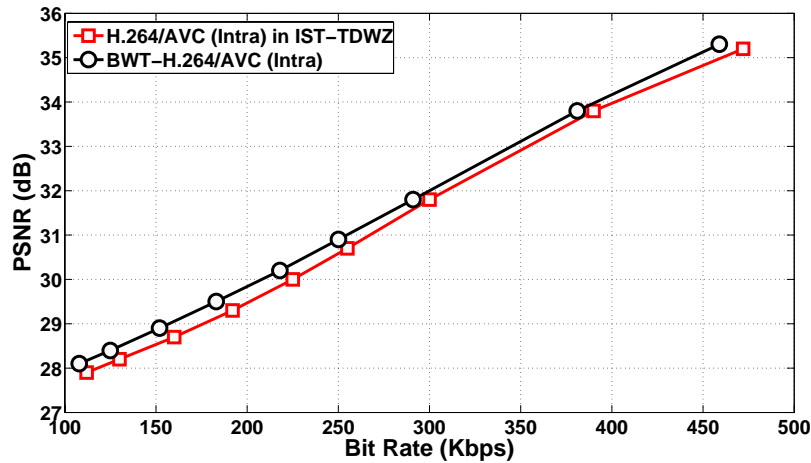
Figure 2.8: Rate distortion performance of key-frames (*Coastguard* at 15 fps)

proposed codec has an average PSNR (dB) gain of 0.47 and 0.4 for *Foreman* and *Carphone* sequences respectively. The RD performances shown here are only for key frames and not the overall performance.
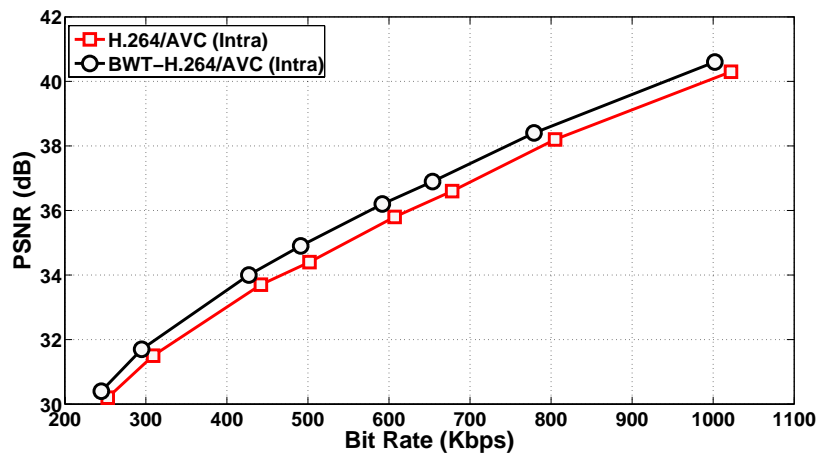
Figure 2.9: Rate distortion performance of key-frames (*Foreman* at 30 fps)

Experiment 3: Analysis of overall rate distortion performance

To evaluate the overall RD performance of the DVC codec, both WZ and key frames are considered. The comparative analysis are shown in Figures 2.11 - 2.18. The GOP size is considered as 2 in the experiments. Following observations are derived.

(a) At 15 fps, the proposed BWT-H.264/AVC (Intra) scheme has better RD
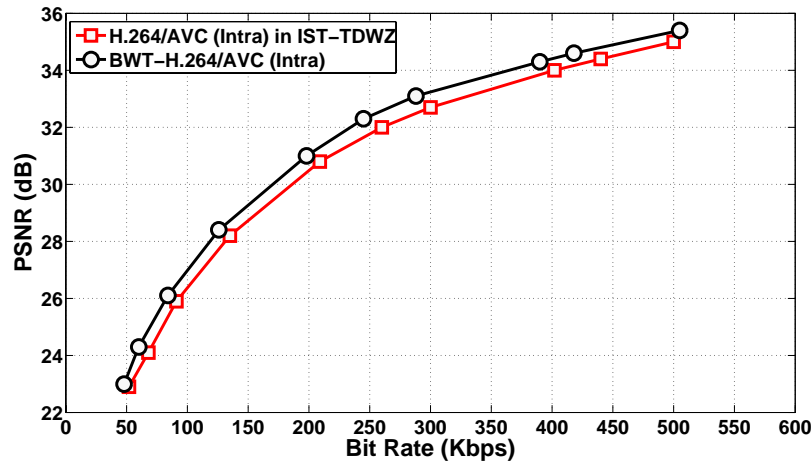
Figure 2.10: Rate distortion performance of key-frames (*Carphone* at 30 fps)

performance for all sequences. Figure 2.11 reflects that proposed codec has
an average PSNR gain of 0.46 dB at lower bit rates and 0.412 dB at higher
bit rates as compared to H.264/AVC (no motion) in *Foreman* sequence which
has inconsistent motion behavior among frames.
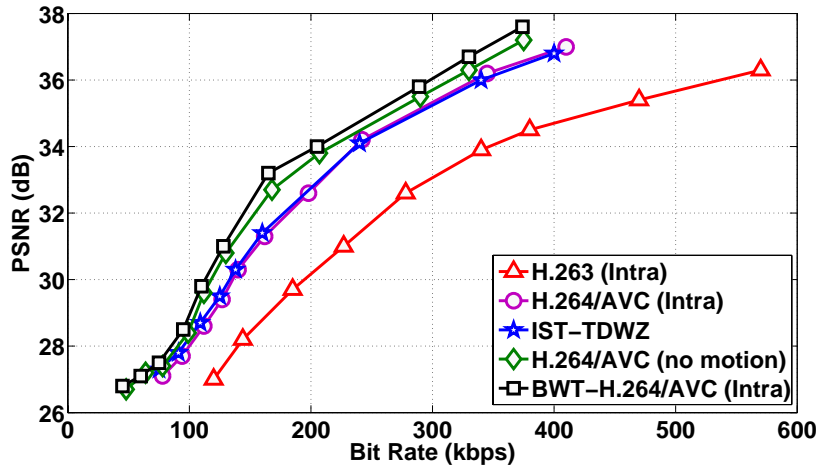
Figure 2.11: Overall rate distortion performance (*Foreman* at 15 fps)

(b) At 15 fps, with uniform and well behaved motion characteristics for
    Coastguard sequence, the proposed BWT-H.264/AVC (Intra) scheme
    manages to defeat the IST-TDWZ scheme with an average PSNR gain of
    0.334 dB. The RD plot for *Coastguard* sequence is shown in Figure 2.12.

(c) *Miss America*, *Carphone* and *Silent* having slow motion characteristics at 15
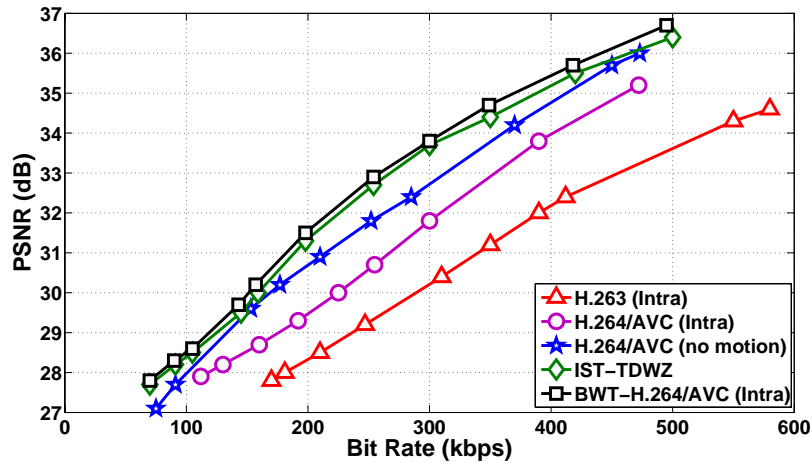
Figure 2.12: Overall rate distortion performance (*Coastguard* at 15 fps)

fps. The proposed codec has an average PSNR gain of 0.346 at low bit rates and 0.421 dB at higher bit rates for *Miss America* sequence over H.264/AVC (no motion). Further, it has an average PSNR gain of 0.398 dB and 0.41 dB over IST-TDWZ codec for *Carphone* and *Silent* sequences respectively.
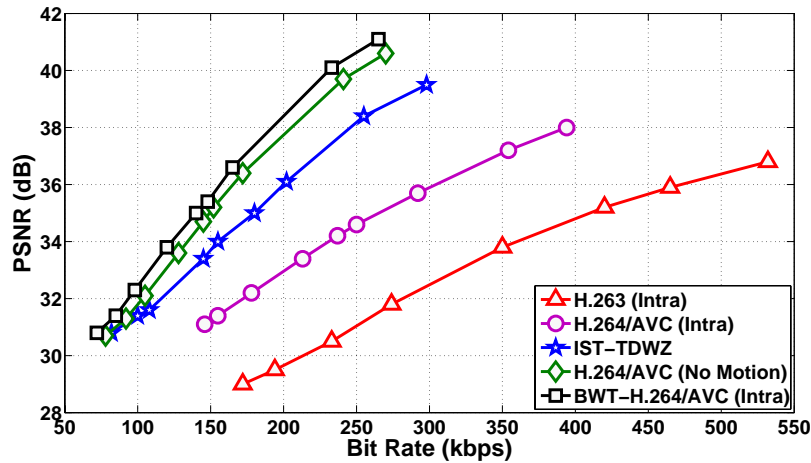
Figure 2.13: Overall rate distortion performance (*Miss America* at 15 fps)

(d) At 15 fps, the proposed BWT-H.264/AVC (Intra) scheme manages to defeat the H.263 (Intra), H.264/AVC (Intra), H.264/AVC (no motion) and IST-TDWZ.

(e) At 30 fps, the frames are more closer in time and thus exists less complex motion between them. The RD plots at 30 fps for different video sequences
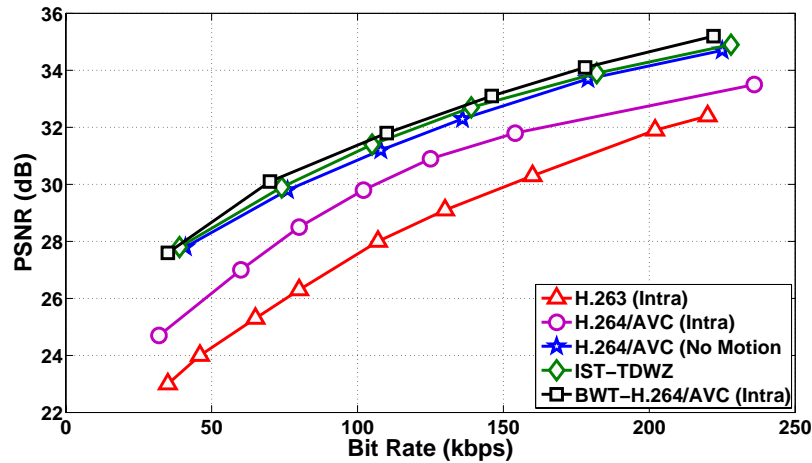
Figure 2.14: Overall rate distortion performance (*Carphone* at 15 fps)
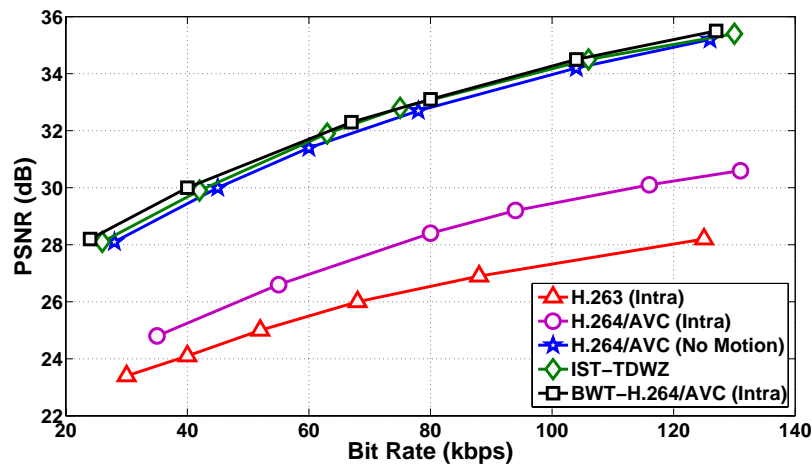
Figure 2.15: Overall rate distortion performance (*Silent* at 15 fps)

are shown in Figures 2.16 - 2.18. The proposed codec outperforms competitive codec schemes for the *Foreman, Carphone* and *Silent* sequences. From the plots it has been observed that the average PSNR gains for *Foreman, Carphone* and *Silent* sequences are 0.523 dB, 0.489 dB, and 0.517 dB over IST-TDWZ codec respectively.

In general, it may be inferred that the proposed BWT-H.264/AVC (Intra) scheme generates improved quality decoded key frames to be used for SI generation. However, for video sequences with 15 and 30 fps, the scheme has a significant PSNR gain as compared to existing schemes.
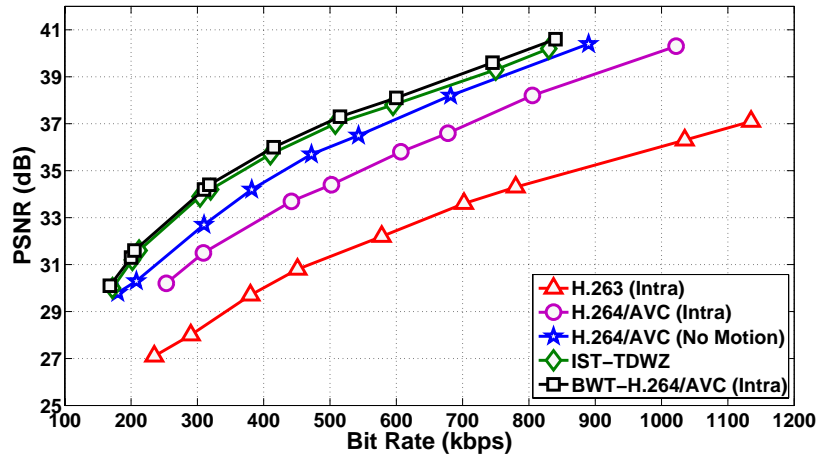
Figure 2.16: Overall rate distortion performance (*Foreman* at 30 fps)
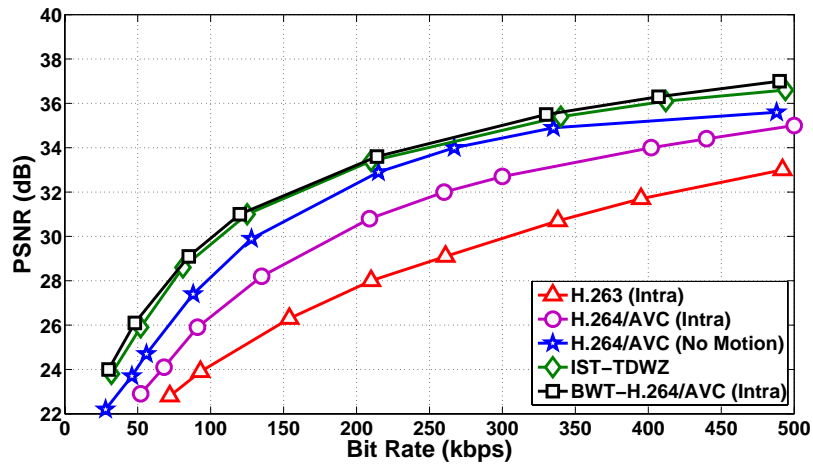
Figure 2.17: Overall rate distortion performance (*Carphone* at 30fps)

Table 2.2 provides an average PSNR gain of the proposed approach over IST-TDWZ codec for different sequences at different bit rates. It can be observed that the average PSNR gain increases along with bit rate.
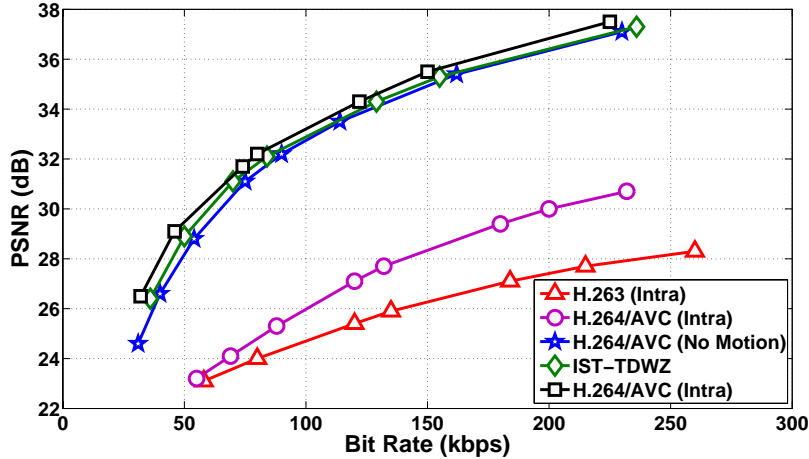
Figure 2.18: Overall rate distortion performance (*Silent* at 30 fps)

Table 2.2: PSNR (dB) gain of BWT-H.264/AVC (Intra) over IST-TDWZ at different bit rate in different video sequence

Sequences (15 fps)	bit rate				
	100	200	300	400	500
Foreman	0.346	0.372	0.455	0.439	0.407
Carphone	0.278	0.324	0.357	0.36	0.382
Coastguard	0.219	0.247	0.294	0.307	0.323
Miss America	0.343	0.36	0.431	0.44	0.45

Experiment 4: Analysis of temporal evaluation (Frame vs. PSNR)

Temporal evaluation is the overall PSNR improvement of the frames of a video sequence after getting reconstructed at decoder end in a DVC. To study this characteristic, PSNR of each frame of *Foreman* and *Coastguard* sequences are ensembled using the proposed scheme and IST-TDWZ. The comparative analysis for both the sequences are shown in Figures 2.19 - 2.21. The schemes consider a GOP size 2 and quantization level Q_8 as defined in [91].

For the *Foreman* sequence shown in Figure 2.19, it is seen that highest PSNR gain of 0.582 dB is obtained for frame number 36 as compared to IST-TDWZ scheme. In general, a gain in PSNR is seen throughout the sequence. For *Coastguard* sequence (Figure 2.20), the PSNR gain is about 0.429 dB for frame number 67. Similarly, a 0.477 dB PSNR gain is seen in *Miss America* and that of 0.532 dB in *Carphone* (Figure 2.21) for some frames.
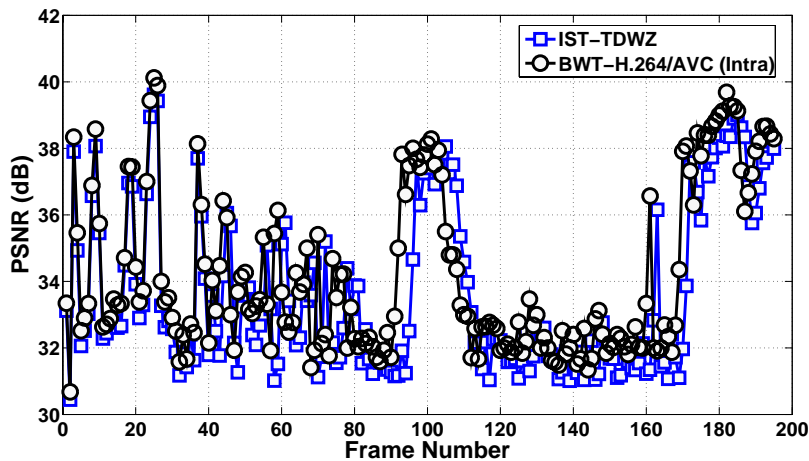
Figure 2.19: Temporal evaluation (*Foreman* at 30 fps)
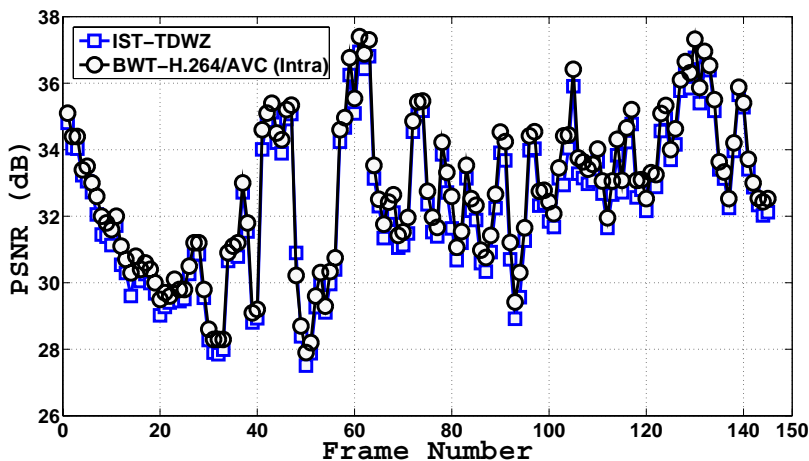
Figure 2.20: Temporal evaluation (*Coastguard* at 30 fps)

Table 2.3 shows the average PSNR for all the decoded frames. For computing the average PSNR both key frames and WZ frames are taken into consideration. The proposed codec has an average gain of 0.41 dB, 0.31 dB, and 0.29 dB over IST-TDWZ scheme for *Foreman*, *Coastguard*, and *Carphone* sequences respectively.

Experiment 5: Analysis of decoding time requirement

To compare the decoding time requirement with latest DVC standard IST-TDWZ, the decoding time (in seconds) are noted in each case and listed in Table 2.4. A GOP size 2 and quantization matrices Q_8 is used uniformly for both schemes. It is observed that our BWT-H.264/AVC (Intra) adds no significant overhead as
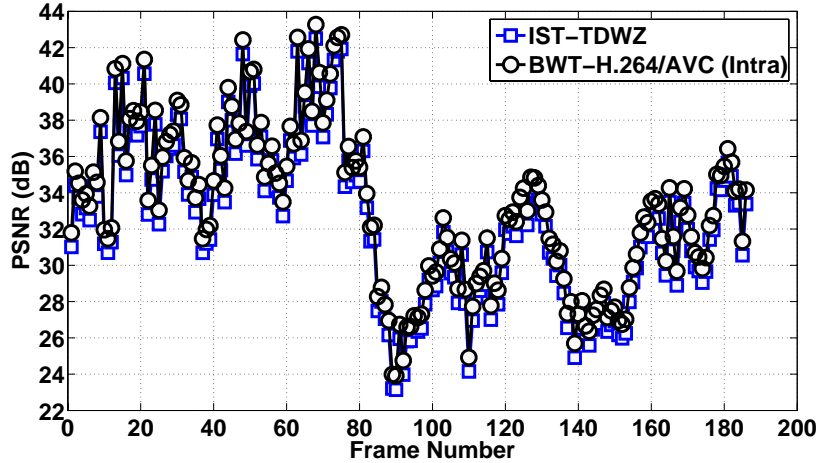
Figure 2.21: Temporal evaluation (*Carphone* at 15 fps)

Table 2.3: PSNR comparison of BWT-H.264/AVC (Intra) for decoded frames

Q_i	Foreman		Coastguard		Carphone	
	IST-TDWZ	BWT-H.264/ AVC(Intra)	IST-TDWZ	BWT-H.264/ AVC(Intra)	IST-TDWZ	BWT-H.264/ AVC(Intra)
1	27.94	28.07	25.81	26.04	24.23	24.44
2	28.92	29.13	26.42	26.69	25.42	25.56
3	29.26	29.34	28.33	28.48	26.31	26.57
4	32.04	32.26	29.28	29.52	27.44	27.64
5	32.13	32.38	30.72	30.98	28.72	28.93
6	33.38	33.54	32.24	32.55	30.61	30.81
7	35.56	35.73	34.21	34.59	32.74	33.11
8	37.42	37.83	36.27	36.58	34.71	35.07

compared to IST-TDWZ scheme, rather our scheme shows a reduction in decoding time due to generation of high quality key frames and WZ frames.

2.5   Summary

This chapter presents a scheme which introduces BWT to the existing H.264/AVC intra-frame coding. BWT attempts to bring regularization to the bit stream prior to compression. The proposed scheme is validated with respect to rate distortion, analysis of bits requirement per frame, PSNR of decoded key frames, PSNR of all frames, decoding time requirement. It has been observed that the proposed BWT-H.264/AVC (Intra) scheme is a better alternative in intra-frame

Table 2.4: Decoding time comparison of BWT-H.264/AVC (Intra)

Sequences	Total decoding time (in seconds)	
	IST-TDWZ	BWT-H.264/AVC (Intra)
Foreman (Q_1)	510.34	487.32
Coastguard (Q_1)	297.12	224.87
Miss America (Q_1)	232.12	204.44
Foreman (Q_2)	637.62	601.12
Coastguard (Q_2)	410.17	374.12
Miss America (Q_2)	308.43	275.18
Foreman (Q_3)	711.78	683.42
Coastguard (Q_3)	477.22	414.31
Miss America (Q_3)	232.12	204.44
Foreman (Q_4)	1124	1107
Miss America (Q_4)	422	373
Coastguard (Q_4)	717	679
Foreman (Q_5)	1292.31	1207.48
Coastguard (Q_5)	822.37	797.54
Miss America (Q_5)	593.44	509.10
Foreman (Q_6)	1804.78	1767.89
Coastguard (Q_6)	1259.39	1193.27
Miss America (Q_6)	834.21	777.27
Foreman (Q_7)	2334.53	2293.44
Coastguard (Q_7)	1670.42	1606.84
Miss America (Q_7)	1015.81	1013.46
Foreman (Q_8)	3698	3613
Miss America (Q_8)	1207	1173
Coastguard (Q_8)	3013	2982

coding framework which has an overall improved performance in all aspects to its competent schemes.

Chapter 3

Dictionary based H.264/AVC intra-frame video coding using OMP

The performance of the distributed video coding (DVC) strongly depends on the key frames quality, its RD performance and the accuracy of side information (SI) generation process. The SI generation estimates the Wyner-Ziv (WZ) frame from preceding and succeeding decoded key frames. Hence improved decoded key frames results to superior SI generation. Normally in DVC, the key frames are coded using intra-frame video coding. This chapter presents a dictionary based intra-frame video coding in DVC with adaptive construction of over complete dictionary. The training is made offline using K-singular value decomposition (K-SVD). K-SVD is an iterative method that alternates between sparse coding of a given signal based on the current dictionary and process of updating the dictionary elements to better fit the data. The dictionary is trained by the residual intra coded macroblock of size 4×4 selected from different video sequences with different motion characteristics. For encoding the dictionary elements, orthogonal matching pursuit (OMP) algorithm has been employed. The popular transform based intra-frame video coding is replaced by a dictionary based approach. The proposed scheme is integrated into H.264/AVC intra coding framework. Performance comparisons have been made with respect to RD performance, peak signal to noise ratio (PSNR), number of requests per frame, decoding time analysis etc. In general, it is observed that the proposed scheme has superior performance as compared to its competitive schemes.

The rest of the chapter is organized as follows. The background of sparse coding and K-SVD algorithm are discussed in Section 3.1. In Section 3.2 the proposed dictionary based H.264/AVC intra-frame coding using OMP (DBOMP-H.264/AVC (Intra)) is given in detail. Section 3.3 deals with simulation results and discussion. Finally, summary of the chapter is given in Section 3.4.

3.1 Background of sparse coding and K-SVD algorithm

In the proposed dictionary based intra-frame coding scheme, we update the dictionary using K-SVD algorithm and subsequently each dictionary element is encoded using OMP algorithm. Here K-SVD utilizes sparse coding technique. For completeness we introduce the theoretical background of sparse coding and K-SVD in sequence.

Recently, sparse coding in a redundant basis has attracted considerable interest in many areas of signal processing applications like compression, denoising, time-frequency analysis, indexing, compressed sensing, audio source separation etc. The basic problem in sparse coding is to represent a given signal as a linear combination of the few signal components from a redundant signal set either exactly or with some acceptable error [106, 107]. For example, from a set of N signal vectors arranged as the columns of a matrix A. Each vector has dimension K where, $K < N$. Using this terminology in sparse approximation theory, we will refer to these vectors *(N)* as atoms and to A as dictionary matrix. So given a signal b, the problem is to identify the fewest atoms whose linear sum will represent b. Mathematically, this can be formulated as solving a system of linear equations given by,

$$AX = b \qquad (3.1)$$

such that X represents the minimum number of non-zero elements, A is the dictionary and b is the given signal. The sparse representation problem is thus formulated as,

$$min\{||X||_0 : AX = b\} \qquad (3.2)$$

where, l_0 norm denotes the number of non-zero elements. The sparse

approximation problem that allows some approximation error is represented as,

$$min\{||X||_0 : ||AX - b||_2 \leq \delta\}, \ \delta \geq 0. \tag{3.3}$$

Matching Pursuit (MP) is a well known algorithm in sparse coding [106, 108, 109]. It is an iterative greedy based approach that selects the atom having the best correlation with the residual vector at each iteration.

Let α_i, $1 \leq i \leq N$ denote the ith atom with $||\alpha_i||_2 = 1$, $\forall \ i$. At the jth iteration, j=1,2,...., the algorithm finds α_j^{opt} as,

$$\alpha_j^{opt} = \underset{\alpha_i \in A}{\arg\max} \ \bigg| < r_{j-1}, \ \alpha_i > \bigg| \tag{3.4}$$

where, A denotes the collection of all atoms commonly called as the dictionary, $< \cdot >$ denotes the inner dot product operation, r_{j-1} represents the residual at $(j-1)^{th}$ iteration with r_0=b. The inner dot product $< r_{j-1}, \alpha_i >$ represents the coefficients associated with atom α_j^{opt} and let us denote it as c_j. The algorithm now updates the residual as,

$$r_j = r_{j-1} - c_j \alpha_j^{opt} \tag{3.5}$$

The algorithm terminates if the norm of the residual falls below a desired approximation error bound, else it proceeds to the next iteration. The approximation at the jth iteration is obtained as,

$$b_j = \sum_{k=1}^{j} c_k \alpha_k^{opt} = [\alpha_1 \ \alpha_2 ......\alpha_j] [c_1 c_2 ......c_j]^T \tag{3.6}$$

The MP algorithm is very simple to implement, however, it is sub-optimal in nature. In addition, it suffers from slow convergence and poor sparsity result. Hence, a modified version of MP algorithm, known as orthogonal matching pursuit (OMP) algorithm is employed to overcome this problem [106, 110]. OMP is fast, powerful, and easy to implement.

The atom selection step in OMP is same as MP however, the residual update stage is different. After each iteration, the coefficients of all the selected atoms are updated such that the residual error is orthogonal to the subspace spanned by the selected atoms. Because of this orthogonalization, once an atom is selected, it is never selected again in subsequent iteration.

Like MP, at j^{th} iteration the algorithm first computes α_j^{opt} as,

$$\alpha_j^{opt} = \underset{\alpha_i \in A - \alpha_{j-1}^{opt}}{\arg\max} \left| < r_{j-1}, \ \alpha_i > \right| \tag{3.7}$$

where α_j^{opt} denotes the selected atom at j^{th} iteration, α_{j-1}^{opt} denotes the set of atoms selected upto $(j-1)^{th}$ iteration, $< \cdot >$ denotes the inner dot product operation, r_{j-1} represents the residual at $(j-1)^{th}$ iteration with $r_0 = b$. The inner dot product $< r_{j-1}, \alpha_i >$ represents the coefficients associated with atom α_j^{opt}. Subsequently the algorithm updates the residual as,

$$r_j = r_{j-1} - c_j \alpha_j^{opt} \tag{3.8}$$

Because of the orthogonal projection, the new residual is orthogonal to all the atoms selected previously including the currently chosen atom and as a result the convergence speed gets improved.

The K-SVD algorithm is flexible and works in conjunction with any pursuit algorithm [111]. The K-SVD is highly efficient, due to an effective sparse coding and Gauss-Seidel like accelerated dictionary update method. The detail description of the K-SVD algorithm is given in *Algorithm* 2.

3.2 Proposed dictionary based H.264/AVC (Intra) intra-frame coding using OMP

The dictionary based training method using sparse coding provides a better approximation of a given signal with fewer number of coefficients as compared to transform based method. Sparse coding using MP and OMP algorithm have been successfully applied in image compression field. However, very few attempts have been made in traditional video coding and for DVC the concept of sparse coding is not exploited so far. The H.264/AVC (Intra) codec is used in most of the DVC solutions. In the present scheme, we have proposed a dictionary based intra-frame coding using OMP in DVC case. The reason for using OMP as compared to MP is that, it shows improved sparsity result and has the ability to increase the convergence speed.

Algorithm 2 : K-SVD Algorithm

Task : To find the best dictionary to represent the data samples $\{b_i\}_{i=1}^N$ as sparse compositions, by solving $\min\limits_{A,X}$ $\{\ ||b - AX||_F^2\}$ subject to $||x_i||_0$ $\leq T_0,\ \forall i$.

Step 1: **Initialization** : Set the dictionary matrix $A^{(0)} \in R^{n \times K}$ with l^2 normalised columns. Set J=1.

Step 2 : **Sparse Coding Stage** : Use any pursuit algorithm to compute the representation vector x_i for each example b_i, by approximating the solution of $\min\limits_{x_i}\{||b_i - Ax_i||_2^2\}$ subject to $||x_i||_0 \leq T_0$, i= 1,2,....,N.

Step 3 : **Codebook Update Stage** : For each column k =1,2,...k in $A^{(J-1)}$ update them using the followings.

   (a) Define the group of examples that use this atom, $\omega_k = 1 \leq i \leq n, x_t^k(i) \neq 0\}$.

   (b) Compute the overall representation error matrix, E_k, by
$$E_k = b - \sum_{j \neq k} d_j x_T^j$$

   (c) Restrict E_k by choosing only one column corresponding to ω_k and obtain E_k^R.

Step 4 : Apply SVD decomposition $E_k^R = U \Delta V^T$. Choose the updated dictionary column $\bar{d}_k$ to be the first column of U. Update the coefficient vector x_R^k to be the first column of V multiplied by $\Delta(1, 1)$
where, $\Delta(1, 1)$ is $n \times n$ diagonal non-negative real values called singular values.

Step 5 : Set $J = J$+1 and repeat Step 2 - Step 5 until the residual error < 0.001.

Initially, a set of 4×4 residual macroblocks from different video frames with varied motion characteristics are selected. These residual macroblocks are arranged in a column format in the dictionary. For constructing an over complete dictionary the number of columns are taken very large as compared to the number of rows and each column in the dictionary is termed as an atom. The over complete dictionary is constructed as $m \times n$ matrices, where rows (m) represent the number of pixel in a block and the columns (n) represent the number of dictionary elements. For training those macroblocks, K-SVD algorithm is used and for sparse signal representation OMP algorithm is employed. During each step in the training the

---

Algorithm 3 Proposed DBOMP-H.264/AVC (Intra)

Step 1 : Select the residual macroblock of size 4×4 with different motion characteristics.

Step 2 : Arrange the residual 4×4 macroblocks in a column format in the dictionary using K-SVD dictionary training. To encode the dictionary elements employ the OMP algorithm.

Step 3 : Update the dictionary elements using K-SVD algorithm. Continue updation process till no further change of elements in the dictionary.

Step 4 : Embed the updated dictionary with H.264/AVC (Intra) of the DVC at encoder side for intra-frame coding.

Step 5 : Select the coefficients value and position (c,p) of the dictionary element and transmit those values to the decoder. Employ uniform quantization and entropy coding to the coefficients. Encode the position using fixed length code. To reduce the dictionary size to code in a given block a position copy method is used.

Step 6 : Apply inverse quantization to reconstruct the 4×4 macroblock.

K-SVD algorithm updates the dictionary atoms along with OMP to accelerate the convergence. The same process is continued and after approximately 62 iterations an updated dictionary is constituted which works well for any residual test samples. The updated dictionary is now embedded with H.264/AVC (Intra) in the encoder side of DVC to perform intra-frame coding. The overall steps of the DBOMP-H.264/AVC (Intra) is given in *Algorithm* 3. For completeness the proposed DBOMP-H.264/AVC (Intra) algorithm is discussed below.

Figure 3.1 shows the proposed dictionary based intra-frame video coding. Here, the video frames are divided into number of macroblocks of size 4×4. Then an intra prediction algorithm is used to predict the pixels in a macroblock using the pixels in the available neighboring blocks. Since there are 9-prediction modes are available for each 4×4 block, a 9-mode prediction is used. Figure 3.2 shows the 9-prediction mode for intra-frame video coding.

Next, the predicted 4×4 macroblock is subtracted from current macroblock to generate the residual 4×4 macroblock. The residual 4×4 macroblock of a video frame are fed as the input to the trained dictionary for sparse coding.
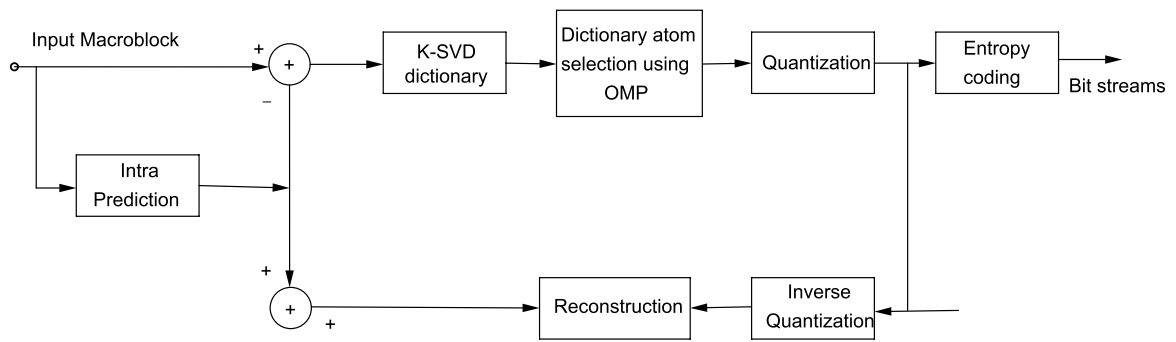
Figure 3.1: Block diagram of proposed DBOMP-H.264/AVC (Intra) scheme
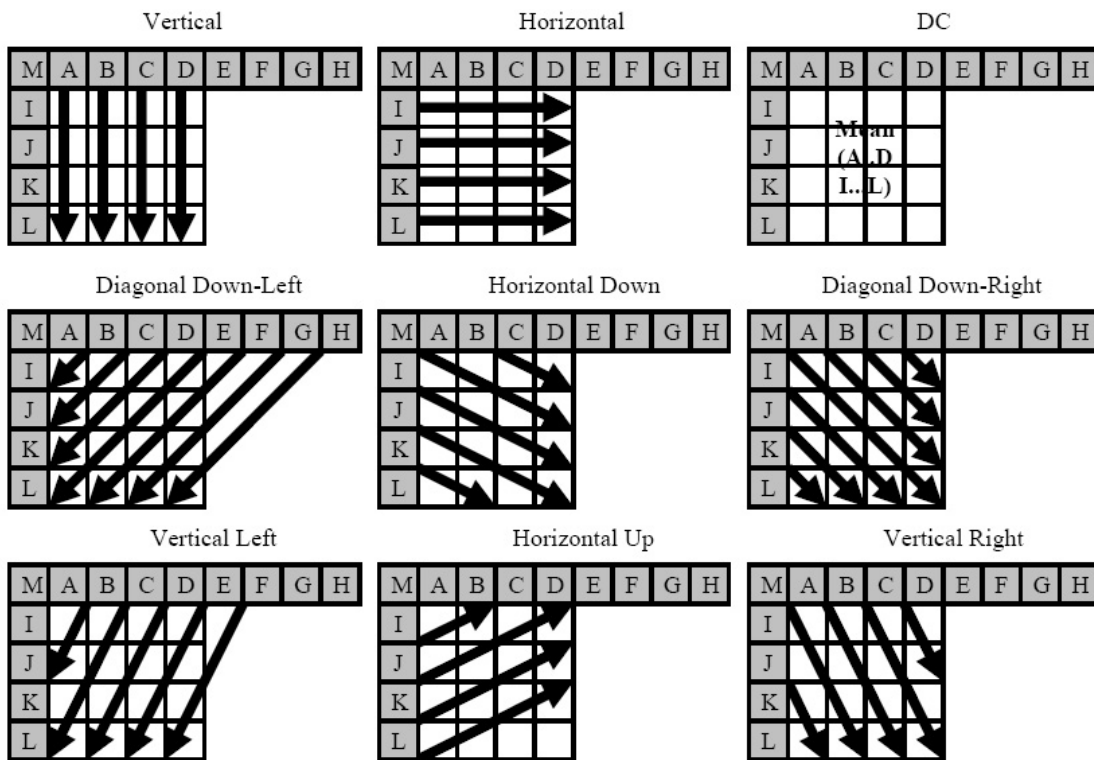
Figure 3.2: Nine mode intra prediction for 4×4 blocks

During the atom selection stage of the sparse coding the OMP algorithm finds the best dictionary elements from the dictionary using inner dot product operation. Next, in the residual update stage, it minimizes the residual error. After the completion of first iteration, the OMP algorithm finds the coefficient (c) value and the corresponding position (p) of the selected atom from the dictionary and send those values to the decoder. The same process is continued and OMP algorithm terminates when the residual error is less than 0.001. The coefficient (c)

is quantized using uniform quantization followed by entropy coding. The position or index (p) is encoded using a fixed length code whose sizes are $\log_2 \lceil m \rceil$, where, m is the number of elements in a dictionary. The dictionary is also available to the decoder. So at the decoder end, the position indicates which dictionary elements to use. Subsequently, inverse quantization is applied to reconstruct the 4×4 macroblock.

To reduce the dictionary size to code in a given block a position copy method is used. Usually in a video frame the neighboring blocks are similar. So to code a current block select the most similar already coded neighboring blocks. A matching criteria known as minimum absolute difference (MAD) is used to indicate which neighboring block is similar to the current block. So instead of sending both position and coefficient value it sends the position of the neighboring block which has minimum cost (i.e. MAD) for the current block. This position copy method is effectively used for homogeneous areas in a video frame where a significant low motion is occurred. For example, block 5 is similar to block 3, so the indices of block 3 is copied and send in place of block 5 as shown in Figure 3.3.
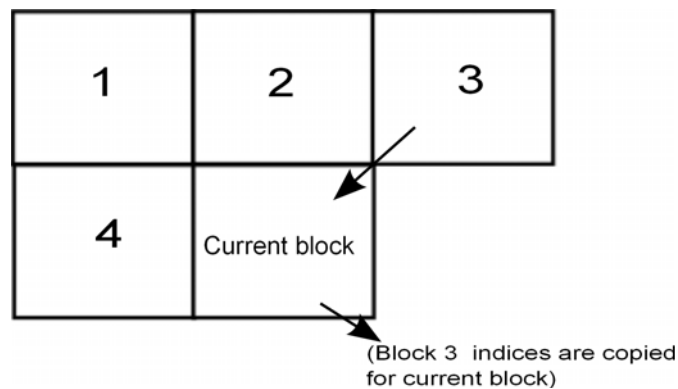
Figure 3.3: Index copy coding method

3.3   Results and discussions

To evaluate the efficacy of the proposed DBOMP-H.264/AVC (Intra) scheme, we integrate it with the Stanford based DVC architecture and various performance measures are computed. Various competitive schemes such as H.263 (Intra), H.264/AVC (Intra), H.264/AVC (no motion) and IST-TDWZ are also simulated

along with the proposed scheme to draw a comparative analysis. The overall study of the proposed scheme has been performed through following simulation experiments.

Experiment 1: Dictionary training methodology

A set of 6700 residual block patches of size 4×4 pixels are taken from different video sequences with different motion characteristics and are considered as training patterns. These blocks constitute 35 frames of *Foreman*, 16 frames of *Miss America*, 17 frames of *Mother and Daughter*, 24 frames of *Container* and 19 frames of *News* video. To reduce the dictionary size, out of 6700 blocks we have selected 470 such blocks which are sorted as per their variance and shown in Figure 3.4. The blocks are arranged in columns of the initial dictionary. So the dictionary size becomes 16×470 where 16 represents the number of rows in the dictionary and 470 represents the corresponding column of the dictionary. The constructed dictionary is trained using K-SVD algorithm along with OMP. To get an updated dictionary, it needs approximately 62 iterations.
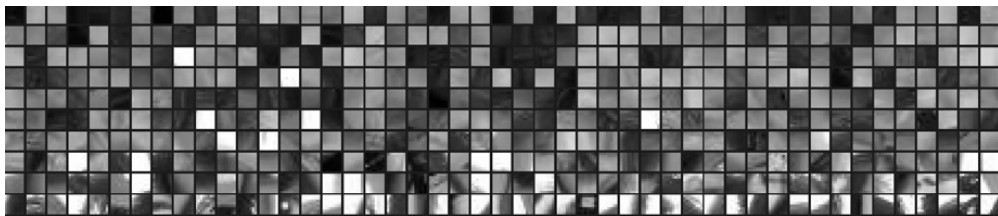
Figure 3.4: A collection of 470 random blocks used in the training, sorted by their variance.

Experiment 2: Analysis of number of bits requirement per frame

The proposed DBOMP-H.264/AVC (Intra) along with DBMP-H.264/AVC (Intra), and traditional H.264/AVC (Intra) used in IST-TDWZ are simulated for *Foreman* and *Coastguard* video sequences using 150 frames from each video. The comparative performance are shown in Figures 3.5 and 3.6 for *Foreman* and *Coastguard* sequences respectively. It may be observed that there is a reduction in bits per frame requirement in both the video sequences, however, it is more significant in *Foreman* video. In particular, it is 8.76 kbits and 6.81 kbits for few frames in *Foreman* and *Coastguard* respectively as compared to DBMP-H.264/AVC (Intra). The reduction in bits per frame are 16.37 kbits and

10.27 kbits over H.264/AVC (Intra) used in IST-TDWZ. In general, the proposed DBOMP-H.264/AVC (Intra) scheme yields a better coding efficiency in terms of bit reduction per frame as compared to its counterparts.

The subjective analysis of the decoded key frames are shown in Figures 3.7 - 3.9 for *Miss America*, *Container*, and *News* sequences respectively. The PSNR (dB) of the decoded key frames for H.264/AVC (Intra), DBMP-H.264/AVC (Intra), and DBOMP-H.264/AVC (Intra) are 26.21, 26.47, and 26.61 respectively for 48th frame number of *Miss America* sequence. Similarly, for 68th frame number of *Container* sequence the PSNR (dB) are 27.31, 27.34, and 27.52 respectively. Figure 3.9 shows the subjective analysis of 89th decoded key frame of *News* sequence. The PSNR (dB) of this frame are 25.23, 25.5, and 26.2 respectively.
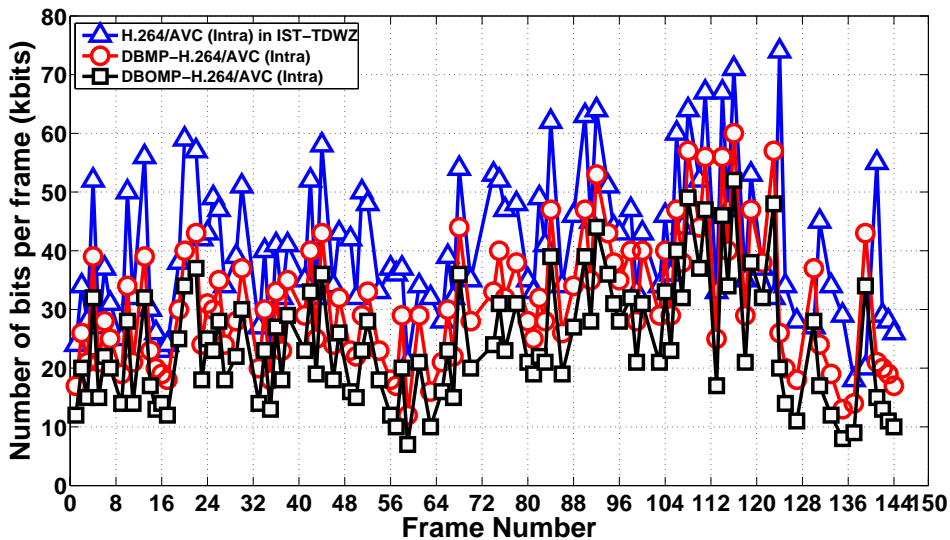
Figure 3.5: Number of bits requirement per frame (*Foreman*)

Experiment 3: Analysis of rate distortion performance of key frames
The RD plots for different video sequences are shown in Figures 3.10 - 3.12 exhibit the RD plots for different video sequences. Following inferences have been drawn from the observations.

(a) At 15 fps, the proposed DBOMP-H.264/AVC (Intra) has better RD performance for *Foreman* and *Coastguard* sequences. In Figure 3.10 for *Foreman* sequence, the proposed scheme has an average PSNR gain of 0.343 dB, and 0.462 dB over DBMP-H.264/AVC (Intra), and H.264/AVC (Intra).
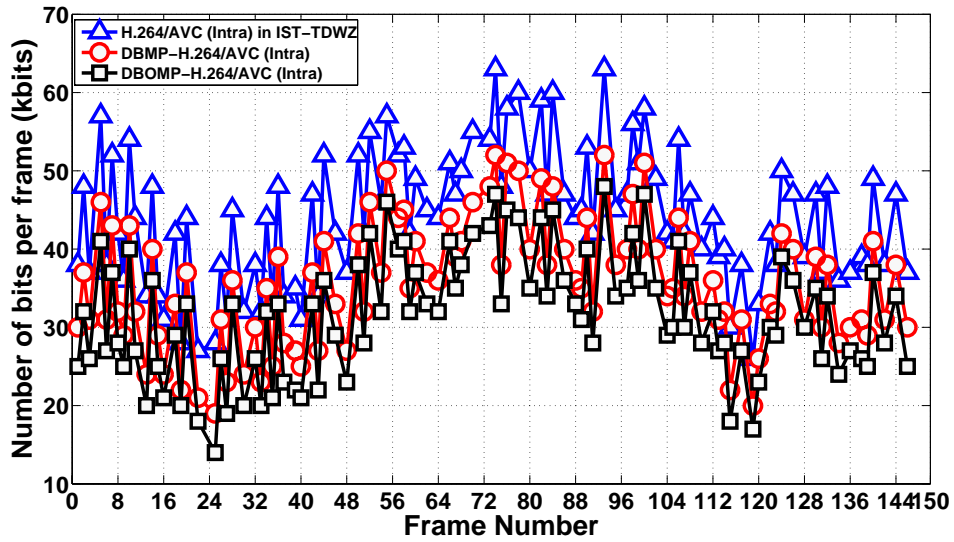
Figure 3.6: Number of bits requirement per frame (*Coastguard*)

(a) H.264/AVC (Intra)

(b) DBMP-H.264/AVC (Intra)

(c) DBOMP-H.264/AVC (Intra)
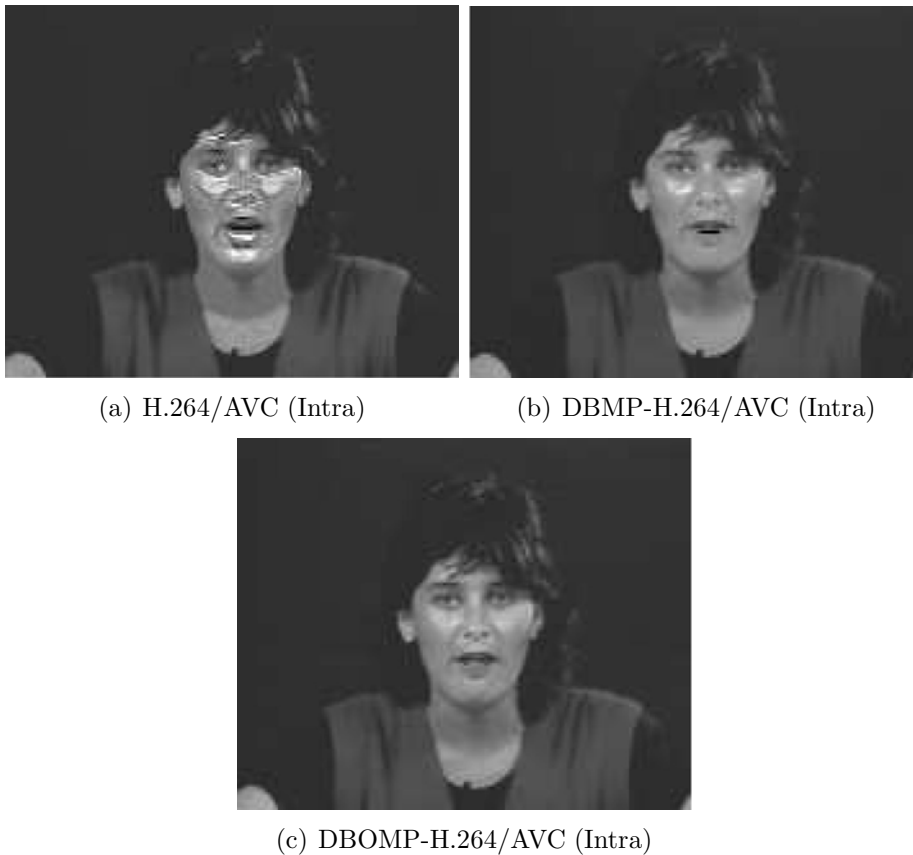
Figure 3.7: Subjective analysis of a decoded key-frame (48th frame of *Miss America*)

(a) H.264/AVC (Intra)   (b) DBMP-H.264/AVC (Intra)

(c) DBOMP-H.264/AVC (Intra)
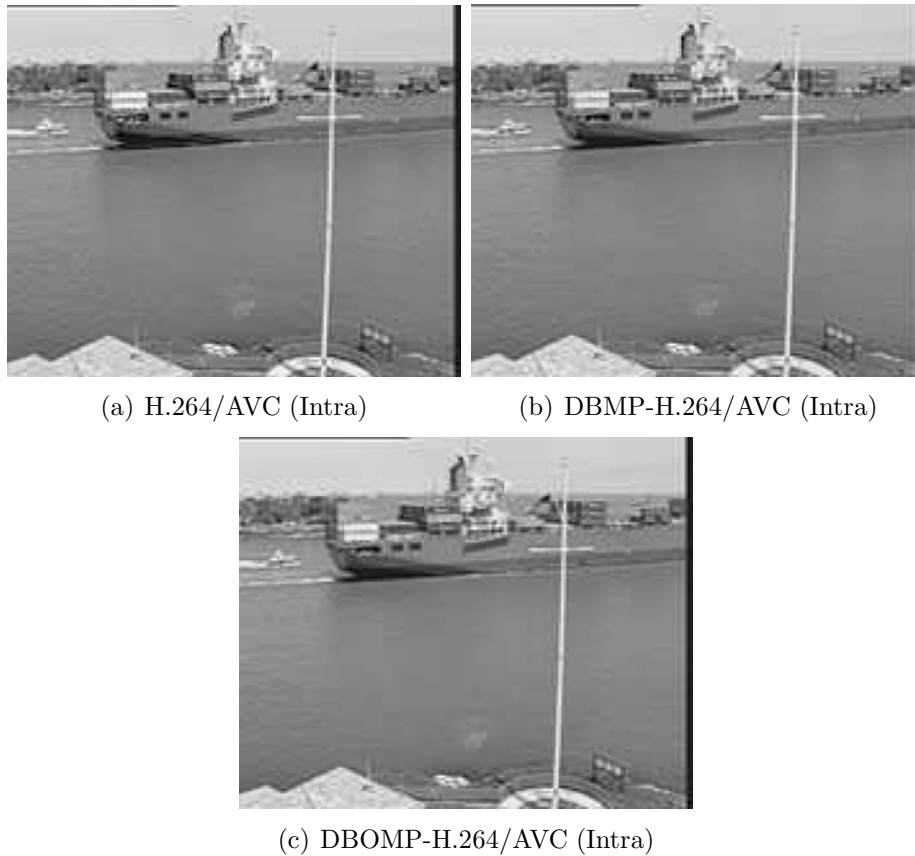
Figure 3.8: Subjective analysis of a decoded key-frame (68th frame of *Container*)

For *Coastguard* sequence shown in Figure 3.11 the average PSNR gain is observed to be 0.21 dB, and 0.327 dB over DBMP-H.264/AVC (Intra) and H.26/AVC (Intra).

(b) Figure 3.12 depicts the RD gain for *Foreman* sequence at 30 fps. The proposed codec has an average PSNR gain of 0.39 dB, and 0.487 dB over DBMP-H.264/AVC (Intra) and H.264/AVC (Intra).

The RD performance shown here are only for key frames and not the overall performance which include both key frames and WZ frames.

Experiment 4: Analysis of overall rate distortion performance

To evaluate the overall RD performance of the DVC codec, both WZ and key frames are used. The comparative analysis are shown in Figures 3.13 - 3.19. The GOP size is considered as 2 in the experiments. Following observations are derived and listed as follows:

(a) H.264/AVC (Intra)   (b) DBMP-H.264/AVC (Intra)

(c) DBOMP-H.264/AVC (Intra)
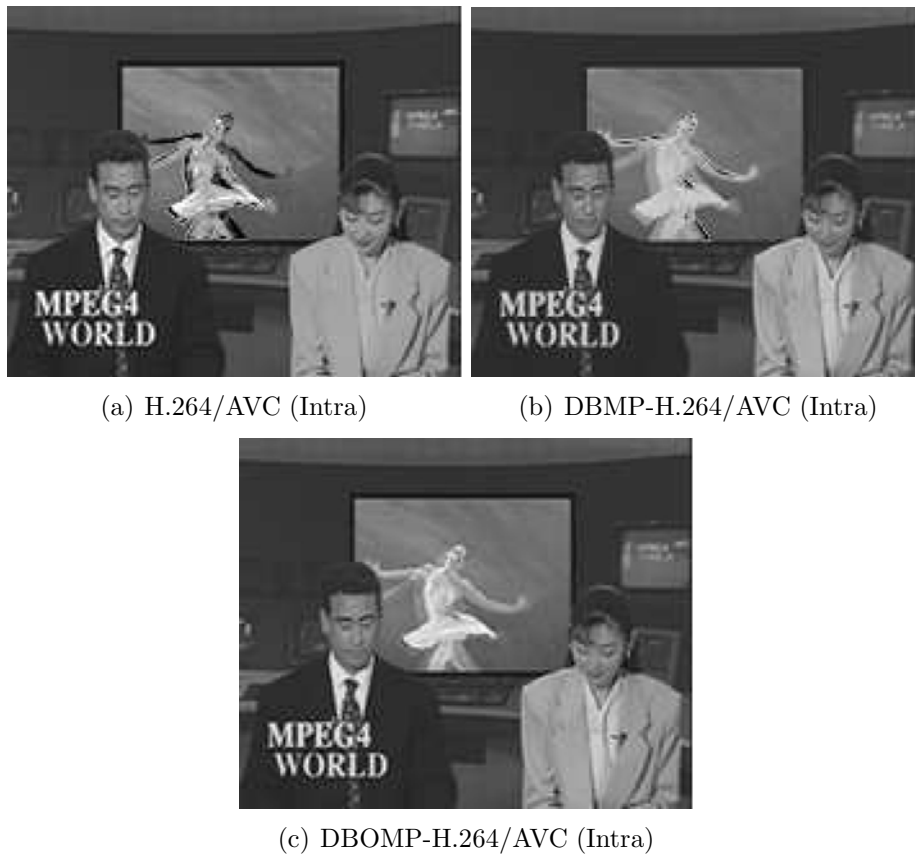
Figure 3.9: Subjective analysis of a decoded key-frame (89th frame of *News*)
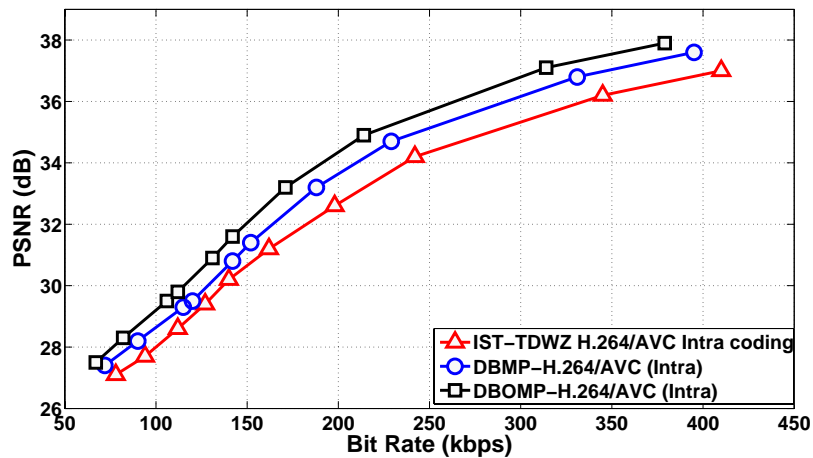
Figure 3.10: Rate distortion performance of key-frames (*Foreman* at 15 fps)

(a) At 15 fps, the proposed scheme has better RD performance for all sequences. Figure 3.13 reflects that the proposed codec has an average PSNR gain of 0.21 dB, 0.34 dB at lower bit rates and 0.31 dB, 0.47 dB at higher bit rates
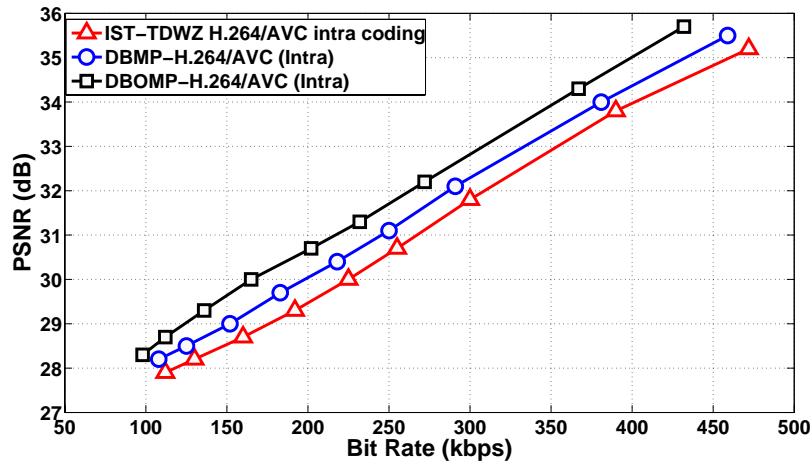
Figure 3.11: Rate distortion performance of key-frames (*Coastguard* at 15 fps)
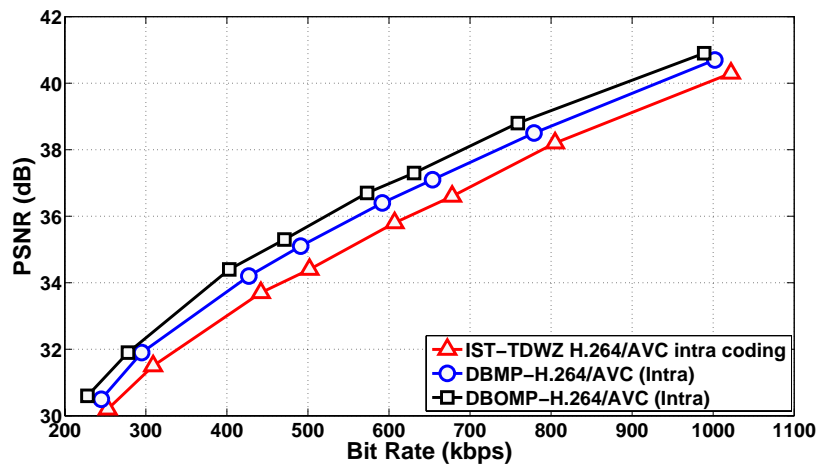
Figure 3.12: Rate distortion performance of key-frames (*Foreman* at 30 fps)

as compared to BWT-H.264/AVC (Intra) and H.264/AVC (no motion) in *Foreman* sequence.

(b) At 15 fps, with uniform and well behaved motion characteristics for *Coastguard* sequence, the proposed codec manages to defeat the BWT-H.264/AVC (Intra) and IST-TDWZ scheme with an average PSNR gain of 0.28 dB and 0.367 dB. The RD plot for *Coastguard* sequence is shown in Figure 3.14.

(c) The proposed codec has an average PSNR gain of 0.23 dB, 0.353 dB at low bit rates and 0.332 dB, 0.437 dB at high bit rates for *Miss America* sequence
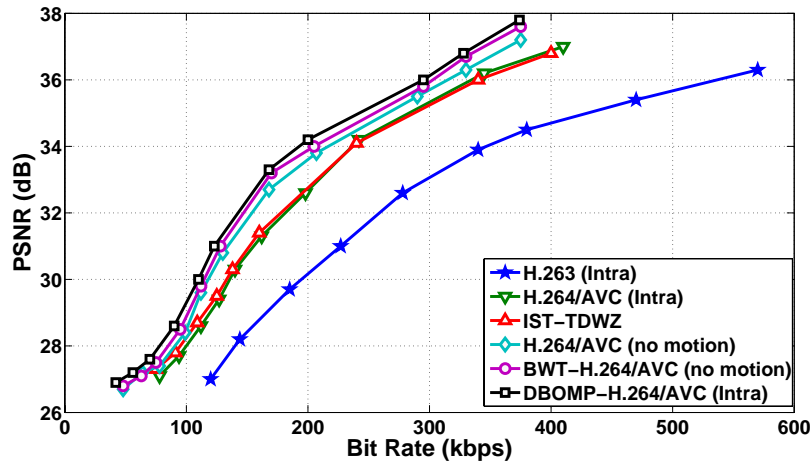
Figure 3.13: Overall rate distortion performance (*Foreman* at 15 fps)
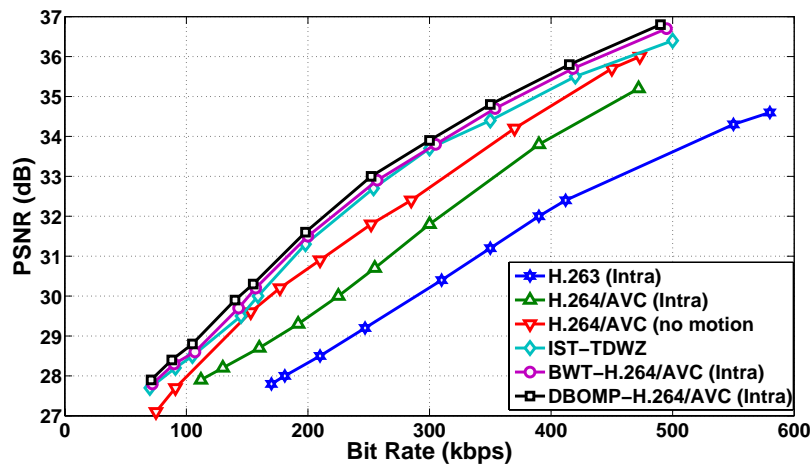
Figure 3.14: Overall rate distortion performance (*Coastguard* at 15 fps)

over BWT-H.264/AVC (Intra) and H.264/AVC (no motion) respectively.

Further, it has an average PSNR gain of 0.31 dB, 0.406 dB and 0.2 dB, 0.25 dB over BWT-H.264/AVC (Intra) and IST-TDWZ codec for *Carphone* and *Silent* sequences respectively.

(d) At 15 fps, the proposed codec manages to defeat the H.263 (Intra), H.264/AVC (Intra), H.264/AVC (no motion), IST-TDWZ and BWT-H.264/AVC (Intra).

(e) The RD plots at 30 fps for different video sequences are shown in Figures 3.18 and 3.19. The proposed codec outperforms competitive codec schemes for
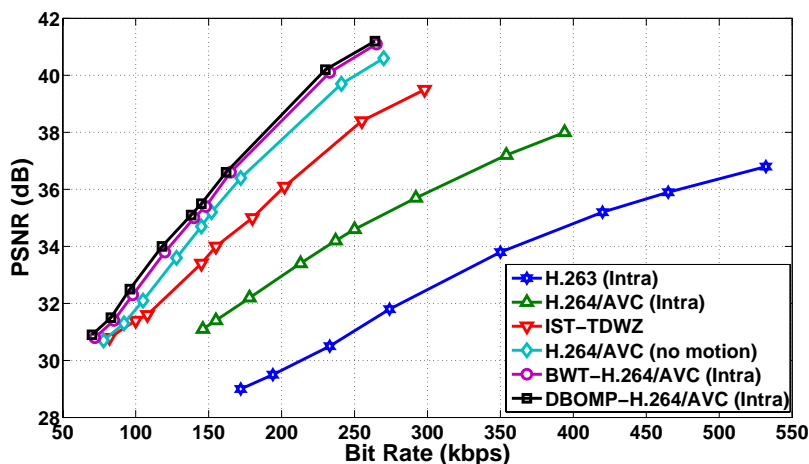
Figure 3.15: Overall rate distortion performance (*Miss America* at 15 fps)
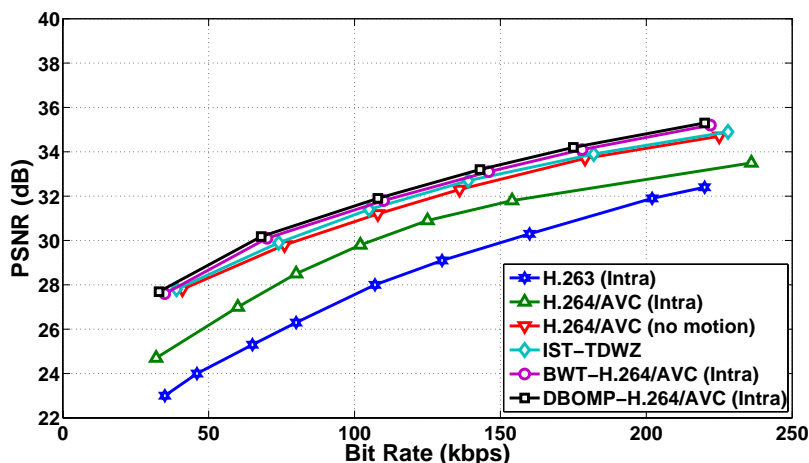
Figure 3.16: Overall rate distortion performance (*Carphone* at 15 fps)

the *Foreman*, and *Carphone* sequences. From the plots it has been observed that the average PSNR gains for *Foreman*, and *Carphone* sequences are 0.44 dB, 0.546 dB and 0.432 dB, 0.492 dB over BWT-H.264/AVC (Intra) and IST-TDWZ codec respectively.

Table 3.1 provides the average PSNR gain of the proposed approach over IST-TDWZ codec for different sequences at different bit rates. It is observed that the PSNR gain increases along with bit rate.

Experiment 5: Analysis of temporal evaluation (Frame vs. PSNR)

This experiment evaluates the temporal evaluation per frame. To study this characteristics, PSNR of each frame of *Foreman* and *Coastguard* sequences are
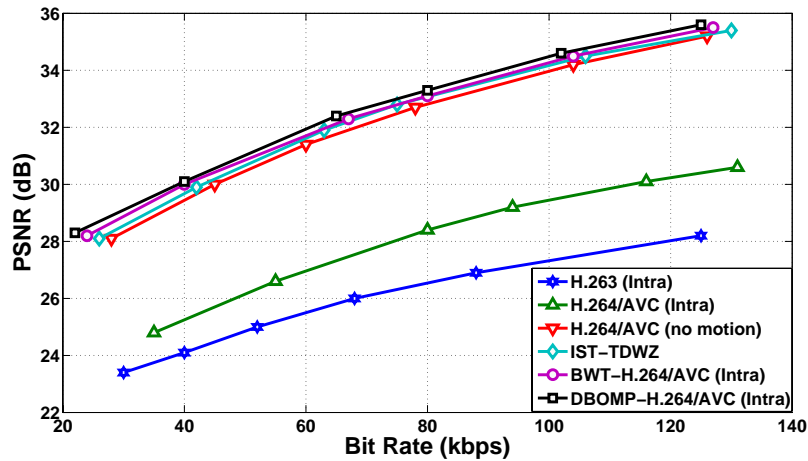
Figure 3.17: Overall rate distortion performance (*Silent* at 15 fps)
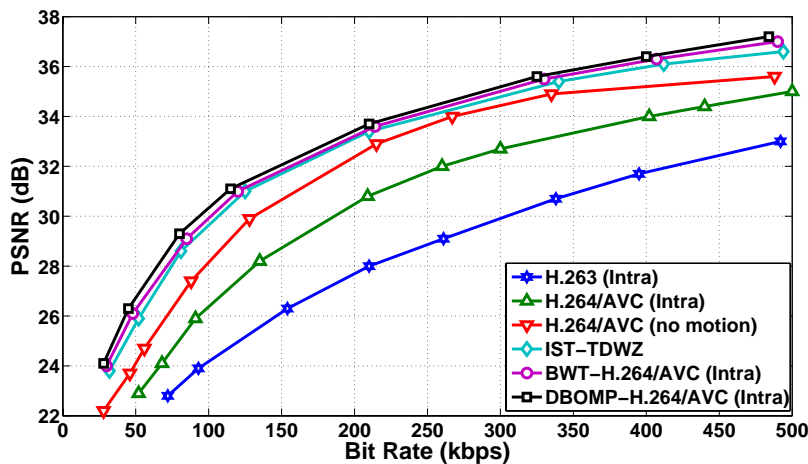
Figure 3.18: Overall rate distortion performance (*Foreman* at 30 fps)

Table 3.1: PSNR (dB) gain of DBOMP-H.264/AVC (Intra) over IST-TDWZ at different bit rate in different video sequence

Sequences (15 fps)	Bit Rate				
	100	200	300	400	500
Foreman	0.363	0.392	0.47	0.476	0.487
Carphone	0.293	0.351	0.364	0.386	0.393
Coastguard	0.232	0.267	0.306	0.324	0.33

computed using the proposed scheme and IST-TDWZ. The comparative analysis for both the sequences are shown in Figures 3.20 and 3.21.

For the *Foreman* sequence shown in Figure 3.20, it is seen the highest PSNR
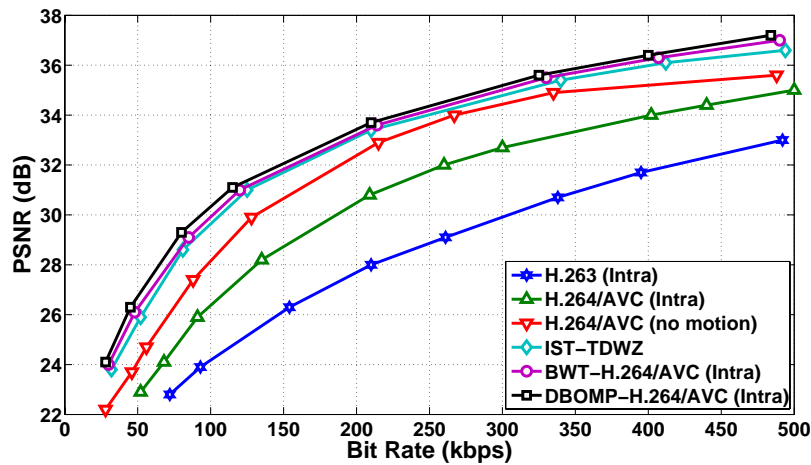
Figure 3.19: Overall rate distortion performance (*Carphone* at 30 fps)

gain of 0.607 dB is obtained for frame number 73. In general, a gain in PSNR is seen throughout the sequence. For *Coastguard* sequence shown in Figure 3.21, the
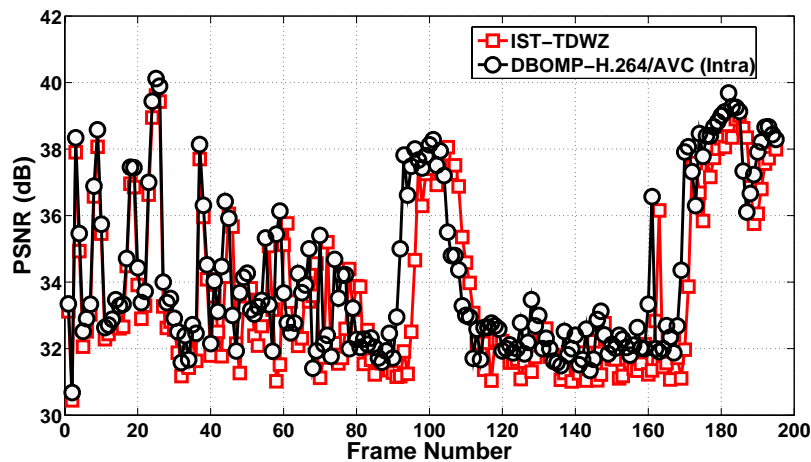
Figure 3.20: Temporal evaluation (*Foreman* at 30 fps)

gain is about 0.446 dB for frame number 69. Similarly, a 0.48 dB PSNR gain is seen in *Miss America* and that of 0.546 dB in *Carphone* for some frames.

Table 3.2 shows the average PSNR for all the decoded frames. The proposed codec has gain of 0.434 dB, 0.32 dB and 0.301 dB over IST-TDWZ scheme for *Foreman*, *Coastguard*, and *Carphone* sequences respectively.
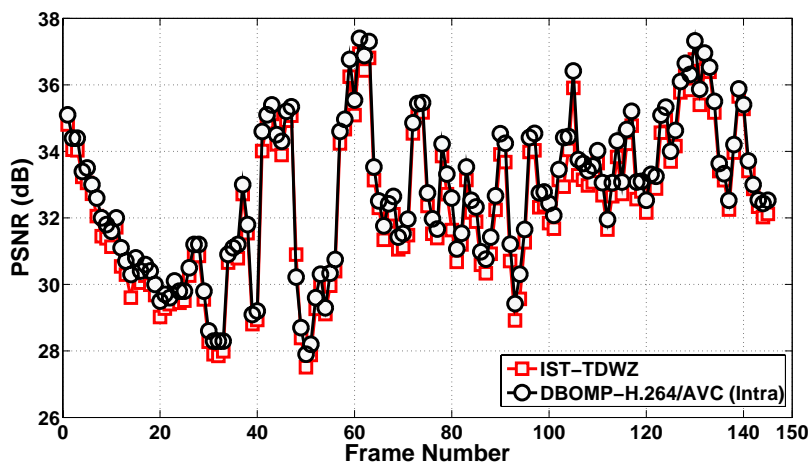
Figure 3.21: Temporal evaluation (*Coastguard* at 30 fps)

Table 3.2: PSNR comparison of DBOMP-H.264/AVC (Intra) for decoded frames

Q_i	*Foreman*		*Coastguard*		*Carphone*	
	IST-TDWZ	DBOMP	IST-TDWZ	DBOMP	IST-TDWZ	DBOMP
1	27.94	29.32	25.81	27.11	24.23	25.56
2	28.92	30.47	26.42	27.63	25.42	26.74
3	29.26	30.68	28.33	29.37	26.31	27.71
4	32.04	33.51	29.28	30.67	27.44	28.49
5	32.13	33.74	30.72	32.04	28.72	30.08
6	33.38	34.39	32.24	33.65	30.61	31.93
7	35.56	36.86	34.21	35.71	32.74	34.17
8	37.42	38.44	36.27	37.33	34.71	36.38

Experiment 6: Analysis of decoding time requirement

To compute the computational time (in seconds), the decoding time for several sequences are computed. For uniformity, an Intel core i3 processor with 3.4 GHZ and 4 GB RAM is used.

Table 3.3 presents the total time required to decode the full sequences with GOP size 2 using 8 quantization matrices as defined in [91]. From the results, it may be noted that there are significant reduction of decoding time for the proposed scheme over IST-TDWZ scheme.

Table 3.3: Decoding time comparison of DBOMP-H.264/AVC (Intra)

Sequences	Total decoding time (in seconds)	
	IST-TDWZ	DBOMP-H.264/AVC (Intra)
Foreman (Q_1)	510.34	463.18
Coastguard (Q_1)	297.12	211.35
Miss America (Q_1)	232.12	191.49
Foreman (Q_2)	637.62	588.62
Coastguard (Q_2)	410.17	353.66
Miss America (Q_2)	308.43	259.38
Foreman (Q_3)	711.78	670.66
Coastguard (Q_3)	477.22	402.11
Miss America (Q_3)	232.12	184.54
Foreman (Q_4)	1124	1084
Miss America (Q_4)	422	357.2
Coastguard (Q_4)	717	666.51
Foreman (Q_5)	1292.31	1185.48
Coastguard (Q_5)	822.37	773.14
Miss America (Q_5)	593.44	492.35
Foreman (Q_6)	1804.78	1733.79
Coastguard (Q_6)	1259.39	1168.44
Miss America (Q_6)	834.21	752.37
Foreman (Q_7)	2334.53	2241.46
Coastguard (Q_7)	1670.42	1589.28
Miss America (Q_7)	1015.81	991.26
Foreman (Q_8)	3698	3579
Miss America (Q_8)	1207	1119
Coastguard (Q_8)	3013	2943

3.4   Summary

This chapter presents an efficient dictionary based intra-frame video coding in DVC with adaptive construction of over-complete dictionary. The training is made off-line and K-SVD is used to train the dictionary. For encoding the dictionary elements, we have employed OMP algorithm. Initially, the residual macroblock of size 4×4 from different video sequences with different motion characteristics are used to train the dictionary. At each iteration, the K-SVD algorithm updates the dictionary elements along with sparse coding. Approximately in 62 iterations, an updated dictionary is resulted which works well for other video sequences that are not used during the training. Different coding results are presented to

show the superiority of the proposed scheme in comparison to its counterparts. Further, the proposed scheme shows overall better performance in terms of both RD performance and subjective perceptual quality.

Chapter 4

An improved side information generation using MLP

Most popular distributed video coding (DVC) solutions use the correlation between original frame with a frame predicted at the decoder. This predicted frame is known as side information (SI), which is a key function in the DVC decoder. The more accurate is the predicted SI, fewer number of bits need to be sent to decode the Wyner-Ziv (WZ) frame. So, SI generation is one of the most focused area of research that directly influence the DVC performance.

This chapter presents a SI generation scheme for distributed video coding based on multilayer perceptron (MLP-SI). The suggested scheme predicts a WZ frame from two decoded key frames. The network is trained offline using patterns from different standard video sequences with varied motion characteristics to achieve generalization. The proposed scheme is simulated along with other standard video coding schemes. Performance comparisons have been made with respect to training convergence, rate distortion (RD), peak signal to noise ratio (PSNR), number of requests per frame, decoding complexity etc. In general, it is observed that the proposed scheme has a superior SI frame generation capability as compared to its competent schemes.

The rest of the chapter is organized as follows. The related literature on SI is described in Section 4.1. Proposed SI generation using MLP (MLP-SI) scheme in DVC is presented in Section 4.2. The results obtained are discussed experiment wise in Section 4.3. Finally, Section 4.4 gives the summary of the chapter.

4.1 Related research on side information generation

The RD performance of DVC strongly depends on the quality of the decoded key frames and the accuracy of the SI generation process. In the original Stanford-based architecture, SI is generated through interpolation of two decoded key frames i.e. preceding and succeeding frames of the original WZ frames. In the recent past, several related works have been reported in the literature for SI generation which involves sophisticated framework. Aaron et al. have proposed two hierarchical frame dependency arrangement in DVC [35]. In their first approach, the SI for the current WZ frame can be extrapolated from a key frame or from a WZ frame. In their second approach, a more complex arrangement has been used with an increase temporal resolution of 2:1 with bidirectional interpolation. With the above two schemes, a poor SI estimation is resulted as group of picture (GOP) size increases and motion becomes more intense and less well behaved. The same authors have proposed solutions using motion compensated interpolation (MC-I) and motion compensated extrapolation (MC-E) [28]. In MC-I, the SI for an even frame at time index t is generated by performing motion compensated interpolation using the two decoded key frames at time $(t-1)$ and $(t+1)$. This interpolation technique involves symmetrical bidirectional block matching for the estimated motion. Since the next key frame is needed for interpolation, the frames have to be decoded out-of-order. This scheme is similar to the decoding of bidirectional frames in predictive coding which is a limitation of this framework. In MC-E, the SI is generated by estimating the motion between the decoded WZ frame at time $(t-2)$ and decoded key frame at time $(t-1)$. Here, the already decoded WZ frames are used for motion estimation. So the reconstruction error from the WZ frame contribute to degradation of SI quality.

Girod et al. have proposed previous extrapolation (Prev-E) and average interpolation (AV-I) techniques to generate SI for low complexity video coding solution [28]. In Prev-E scheme, the previous key frame is used directly as SI whereas, in AV-I technique, the SI for the WZ frame is generated by averaging the pixel values from the key frames at time $(t-1)$ and $(t+1)$. The limitation

of Prev-E scheme is that it does not employ motion compensation to generate SI. So the Prev-E scheme is better than DCT based intra-frame coding only at lower bit rates. In AV-I scheme, the pixel values from key frames $(t-1)$ and $(t+1)$ are averaged and it is not sufficient to generate a good quality SI. A hash based motion compensation is proposed by the same authors in [112].

In 2005, Tagliasaccchi et al. have presented a motion compensated temporal filtering technique [113]. This scheme is based on pixel domain coding solution. Natario et al. have proposed a motion field smoothing algorithm to generate SI in [83]. Artigas and Torres have proposed an iterative motion compensated interpolation technique where the turbo decoder runs several times for decoding the WZ frame to estimate SI and as a result a significant delay is associated [84].

Adikari et al. have proposed a multiple SI stream for DVC [85]. It uses two SI streams which are generated using motion extrapolation and compensation (ME-C). The first SI stream (SS-1) is predicted by extrapolating the motion from the previous two closest key frames. The second SI stream (SS-2) is predicted using the immediate key frame and the closest WZ frame. Fernando et al. have proposed a SI scheme using sequential motion compensation, using both luminance and chrominance information to improve the decoding performance of DVC [86]. This work has been extended by Weerakkody et al. in [87]. Here, a spatio-temporal refinement algorithm is used to improve the SI resulting from motion extrapolation.

In 2006, Kubasov et al. have presented a mesh-based MC-I approach to resolve the problems occurred in the SI interpolation. Their approach is based on block translational motion model [114]. In this scheme, it addresses the problem of motion discontinuities and occlusions. The overall increasing accuracy of SI leads to the improvement in WZ coding. However, this scheme suffers from a high computational complexity burden due to the mesh based structure adopted in the framework. A technique to improve the performance by sending additional hash information to help SI has been presented in [115]. However, this framework is suitable for PRISM based architecture rather than Stanford-based architecture. Tagliasaccchi et al. have used a Kalman filter to generate improved SI at the decoder [116]. This scheme shows better coding performance. However, it

increases the encoder complexity due to motion estimation task being performed at the encoder. Badem et al. have proposed a novel SI refinement technique based on motion estimation in DCT domain for DVC [88]. Varodayan et al. have proposed an unsupervised motion vector learning algorithm for SI generation [89]. This method applies an expectation maximization (EM) algorithm for unsupervised learning of motion vectors. The authors have claimed a better RD performance.

Brites et al. have proposed a frame interpolation framework with forward and bi-directional motion estimation with spatial smoothing [38]. This framework is commonly known as Instituto Superior Tecnico Transform Domain Wyner-Ziv (IST-TDWZ) video coding and used by many DVC researchers. However, the most promising SI refinement framework has been adopted and extended by the same group in their other contribution in [91]. Here, first both the key frames are low-pass filtered. Then a block matching algorithm is used to make the motion estimation between two adjacent key frames. The bi-directional motion estimation is performed followed by the forward motion estimation. Once the final motion vectors are obtained, the bidirectional motion compensation is performed. As so many modules are associated to create SI, it affects the decoding time complexity. The distributed coding for video services (DISCOVER) codec which also uses the same SI generation framework proposed by IST group [90]. IST-TDWZ video codec shows similar performance as DISCOVER video codec. The DISCOVER codec uses lower density parity check (LDPC) code whereas IST group uses turbo code.

It is observed from the literature that SI generation is mostly performed from the decoded key frames information. In this chapter, we propose a similar but efficient scheme based on blockwise pixel estimation to generate a high quality SI frame. The proposed scheme utilizes a neural network for this purpose.

4.2 Proposed MLP based SI generation

The proposed scheme is motivated by the fact that in a video sequence, the pixel locations are shifted frame by frame due to motion. This inter-frame pixel movement is found out to be non-linear due to 3-D motions of an object moving

Table 4.1: Non-linear pixel movement in three consecutive frames of *Foreman* video sequence

$(i-1)^{th}$ frame pixel coordinate	New pixel coordinate in $(i)^{th}$ frame	New pixel coordinate in $(i+1)^{th}$ frame	Euclidean distance of pixel coordinate in $(i-1)^{th}$ frame and $(i)^{th}$ frame	Euclidean distance of pixel coordinate in $(i)^{th}$ frame and $(i+1)^{th}$ frame
(1,2)	(3,1)	(4,2)	2.23	1.414
(1,3)	(1,5)	(2,4)	2	1.414
(1,4)	(3,2)	(2,6)	2.82	4.123
(1,5)	(1,7)	(2,3)	2	7.071
(1,6)	(1,8)	(7,6)	2	6.324
(1,8)	(2,8)	(3,4)	1	4.123
(2,3)	(5,3)	(3,2)	3	2.23
(2,4)	(7,5)	(6,3)	5.099	2.23
(2,6)	(6,7)	(1,7)	4.123	5
(2,8)	(7,6)	(1,4)	5.385	6.324
(3,2)	(5,2)	(4,3)	2	1.414
(3,3)	(3,6)	(3,7)	3	1
(5,1)	(8,1)	(6,8)	3	7.280
(5,3)	(4,4)	(4,5)	1.414	1
(5,4)	(6,3)	(6,4)	1.414	1
(5,6)	(7,3)	(8,3)	3.605	1
(6,1)	(8,2)	(8,5)	2.23	1.73
(6,2)	(7,2)	(1,2)	1	2.449
(6,8)	(8,5)	(1,6)	3.605	7.071
(7,3)	(6,4)	(4,8)	1.414	4.472
(7,7)	(5,8)	(1,8)	2.23	2

back and forth in horizontal, vertical, and diagonal direction or may be due to some directional orientation, camera zooming, and panning. The linear motion and non-linear motion of a pixel in a video between frames is shown in Figure 4.1.
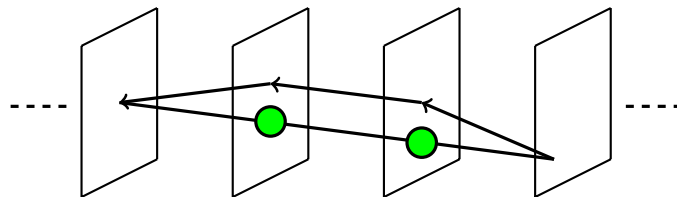
Figure 4.1: Linear vs. non-linear motion between adjacent frames.

To support our assumption of non-linear pixel movement among frames, following experiment is conducted. Three consecutive frames, namely 103,104 and 105 with good motion patterns are considered from *Foreman* video. The shift of various pixel locations in consecutive frames are tabulated in Table 4.1.
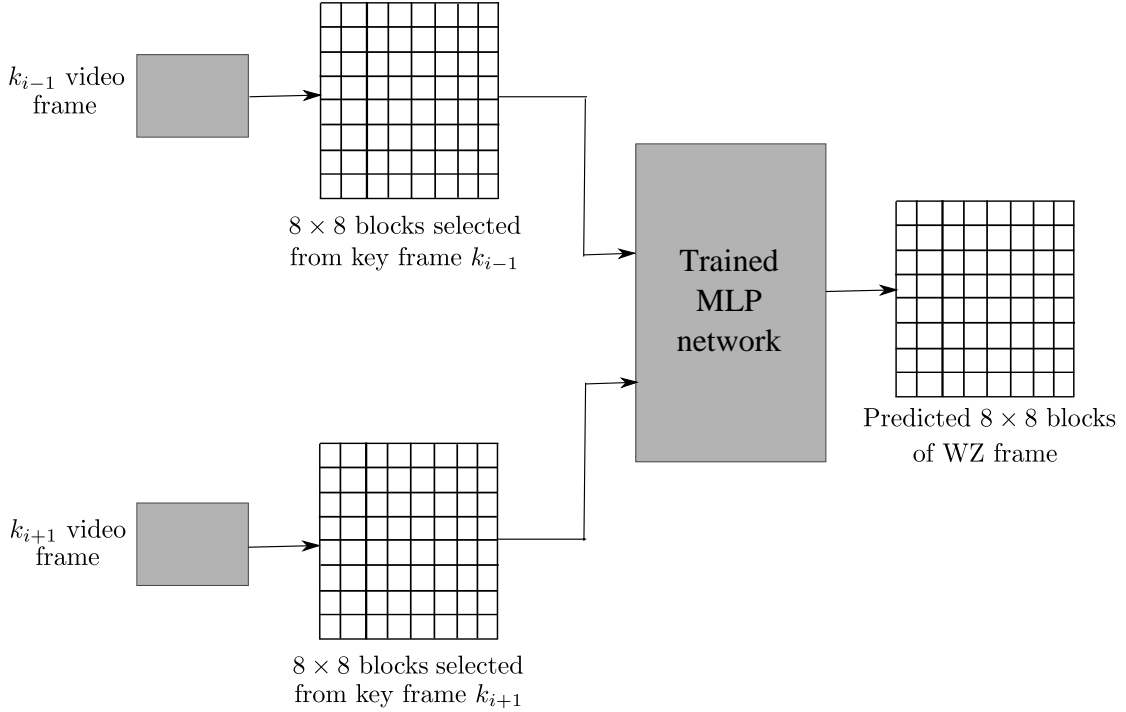
Figure 4.2: Block diagram of neural predictor

The Euclidean distance measures reflect the non-linear motion property of a pixel. It may be observed that the pixel distance between $(i-1)^{th}$ - i^{th} and i^{th} - $(i+1)^{th}$ pixels are different and non-linear. Similar observations are found in pixels with other frames of *Foreman, Coastguard,* and other video sequences. Artificial neural network (ANN) being a potential tool for non-linear prediction is utilized here to predict a 8×8 WZ frame block from given two 8×8 key frame blocks. To generate the training patterns, 8×8 blocks form key frames and WZ frames are collected from different video sequences with various motion patterns i.e. the frames with vertical, horizontal, and diagonal motion patterns are selected. The two 8×8 key frames blocks are used as input and corresponding 8×8 WZ frame block is used as the target for training. The block diagram of the neural predictor is shown in Figure 4.2. The overall WZ frame ensemble from the predicted blocks is used as SI frame. The expanded block diagram is shown through the neural
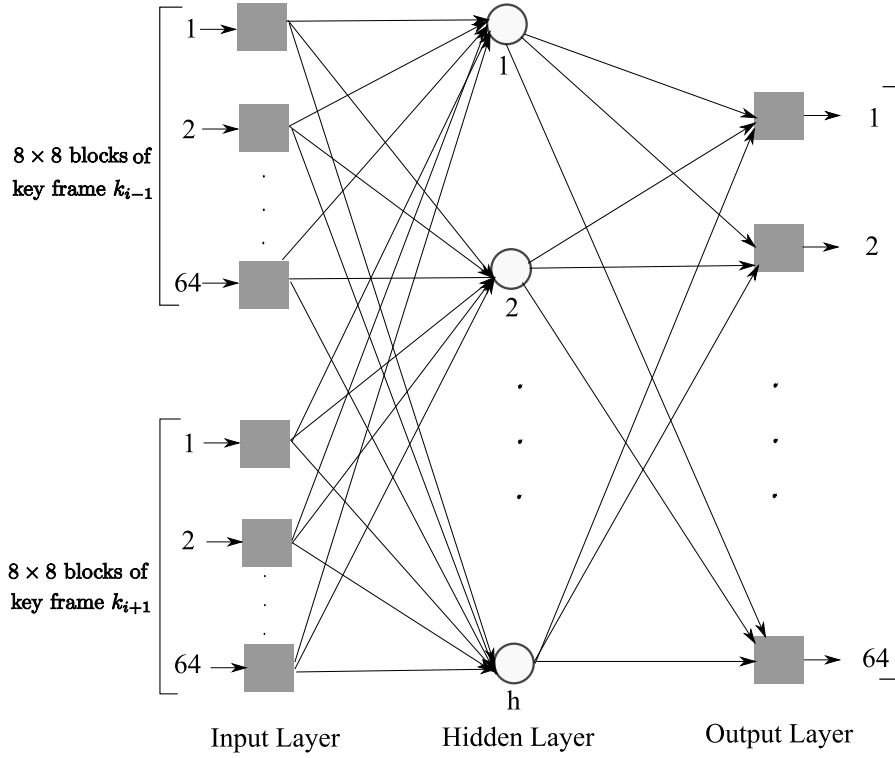
Figure 4.3: Architecture of the neural predictor

architecture in Figure 4.3. *Algorithm* 4 lists the steps of the proposed scheme.

Algorithm 4 : MLP-SI scheme

Training Phase: (Offline)

Inputs : 8×8 blocks of $(i-1)^{th}$ and $(i+1)^{th}$ key frames.

Outputs : 8×8 block of WZ frame (i^{th} frame).

1. Select a representative training set consisting of key frame input blocks of size 8×8 and the corresponding WZ block of size 8×8 as target.

2. Train the network till convergence is achieved with respect to MSE (dB).

3. Fix the number of hidden layer and number of inputs through experimentation for less convergence error.

Working Phase: (Decoder side of DVC)

1. Split the decoded key frames into 8×8 blocks.

2. Pass the input blocks to the trained MLP to predict the 8×8 WZ block.

3. Reconstruct the WZ frame from the estimated blocks.

4.3   Results and discussions

To evaluate the efficacy of the proposed MLP-SI scheme, it is integrated with the Stanford based DVC architecture and various performance measures are computed. Competent schemes such as H.263 (Intra), H.264/AVC (Intra), H.264/AVC (no motion) and IST-TDWZ are also simulated along with the MLP-SI scheme to perform a comparative analysis. The overall simulation study has been divided into six different experiments and they discussed below in detail.

Experiment 1: Convergence characteristics of MLP

To train the MLP, a training set constituting of 8×8 blocks from key frames and corresponding 8×8 from WZ frames are selected randomly to use them as inputs and targets respectively. The frames are chosen from different standard video sequences namely, *Foreman*, *Miss America*, *Mother and Daughter*, *Container*, and *News*. The selected frames include different motion characteristics to constitute a representative training set. A total of 7100 training patterns are chosen and they constitute 45 frames of *Foreman*, 23 frames of *Miss America*, 20 frames of *Mother and Daughter*, 30 frames of *Container*, and 24 frames of *News* video. The number of hidden layers and number of neurons in each layer is decided through experimental study and finally, an ANN architecture of 128-14-64 is selected. The network is trained using scaled conjugate gradient learning algorithm utilizing batch updating procedure. The training convergence characteristics is shown in Figure 4.4. It may be observed that the network converges in approximately 200 iteration. Subsequently, the trained MLP network is used to predict the WZ frame blocks.

Experiment 2: Performance analysis of side information

Among the various competent schemes IST-TDWZ is a recent SI generation scheme in DVC scenario. Hence PSNR (dB) for frames *Foreman* and *Coastguard* sequences are compared with this scheme. The comparative results of PSNR (dB) for *Foreman* and *Coastguard* sequences are shown in Figures 4.5 and 4.6 respectively. It may be observed that the SI frames generated by MLP-SI scheme has a superior PSNR performance as compared to IST-TDWZ. This reflects the qualitative side information of WZ frames by MLP-SI scheme. The subjective
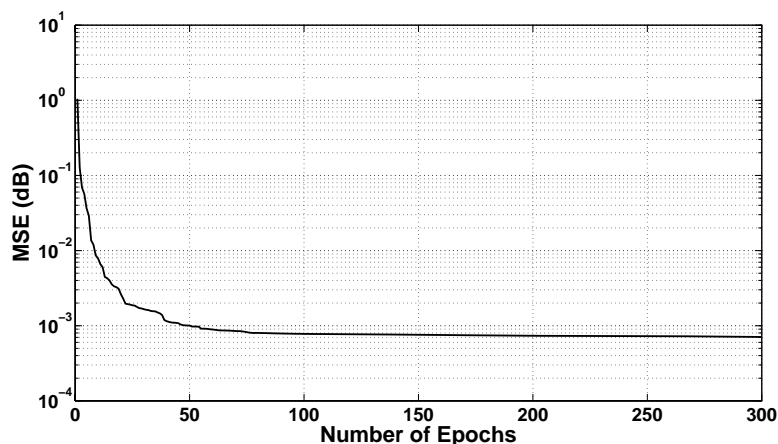
Figure 4.4: Training convergence characteristics of MLP
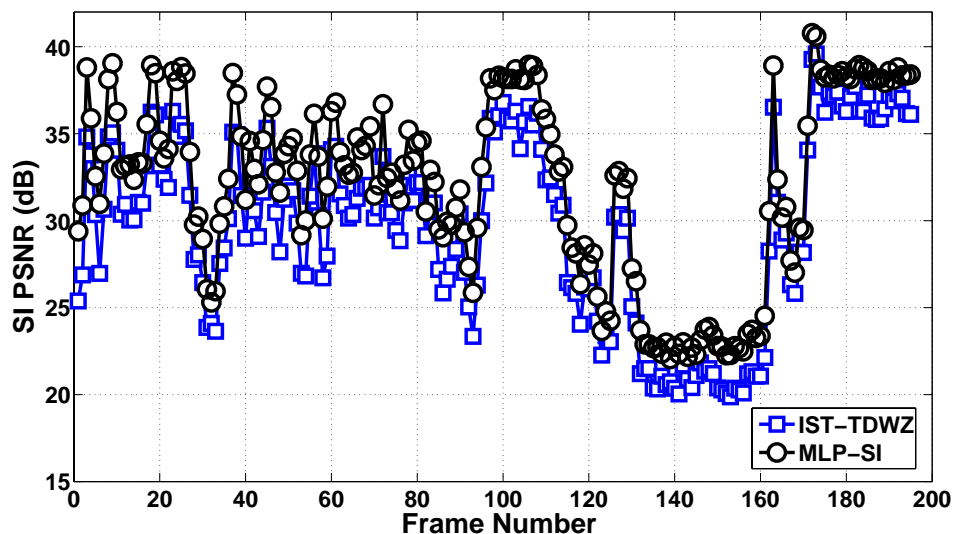
Figure 4.5: PSNR of SI frames (*Foreman* at 15 fps)

performance are shown in Figures 4.7 - 4.9 for *Foreman, Mother and Daughter*, and *Carphone* sequences. It may be noted that the MLP-SI frame is visually more appealing as compared to IST-TDWZ.

Experiment 3: Analysis of overall rate distortion performance

To evaluate the overall efficacy of the proposed MLP-SI generation scheme, it is integrated with the Stanford based DVC architecture. If the SI frame quality is better in any scheme, PSNR (dB) will be higher with less bit rate. This is known as RD performance of a scheme. To demonstrate the RD performance of different standard video codecs along with our MLP-SI scheme, simulation results
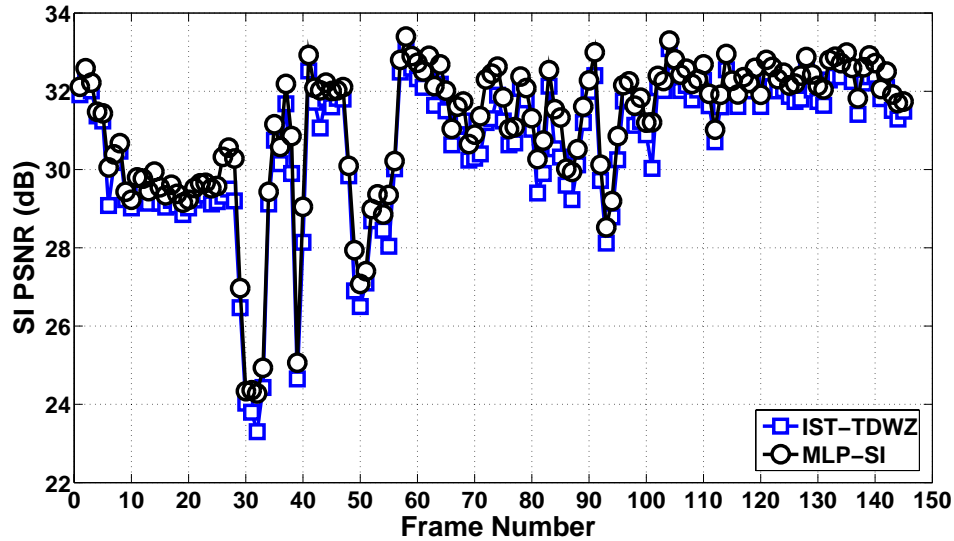
Figure 4.6: PSNR of SI frames (*Coastguard* at 15 fps)

(a) IST-TDWZ (PSNR = 31.2 dB)   (b) MLP-SI (PSNR = 31.53 dB)

Figure 4.7: Subjective analysis of a decoded SI frame (167th frame of *Foreman*)

are shown in Figures 4.10 - 4.17. Following observations are derived,

(a) MLP-SI has an average PSNR gain of 0.52 dB at lower bit rate and 0.44
dB at higher bit rate as compared to H.264/AVC (no motion) in *Foreman*
sequence. At 15 fps, for *Coastguard* sequence with uniform and well behaved
motion, the proposed MLP-SI has an average PSNR gain of 0.432 dB
over IST-TDWZ. Unlike for *Carphone* sequence the average PSNR gain
is 0.487 dB over IST-TDWZ. At 15 fps, with slow motion characteristics,
the proposed MLP-SI scheme has an average PSNR gain of 0.38 dB over
H.264/AVC (no motion) for *Miss America* sequence and 0.468 dB over

(a) IST-TDWZ (PSNR = 28.37 dB)   (b) MLP-SI (PSNR = 28.64 dB)

Figure 4.8: Subjective analysis of a decoded SI frame (178th frame of *Mother and Daughter*)

(a) IST-TDWZ (PSNR = 26.12 dB)   (b) MLP-SI (PSNR = 26.4 dB)

Figure 4.9: Subjective analysis of a decoded SI frame (102th frame of *Carphone*)
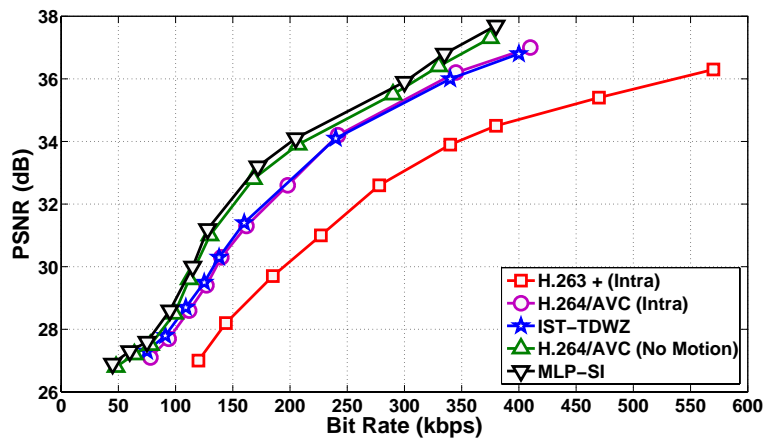
Figure 4.10: Overall rate distortion performance (*Foreman* at 15 fps)

IST-TDWZ for *Silent* sequence.
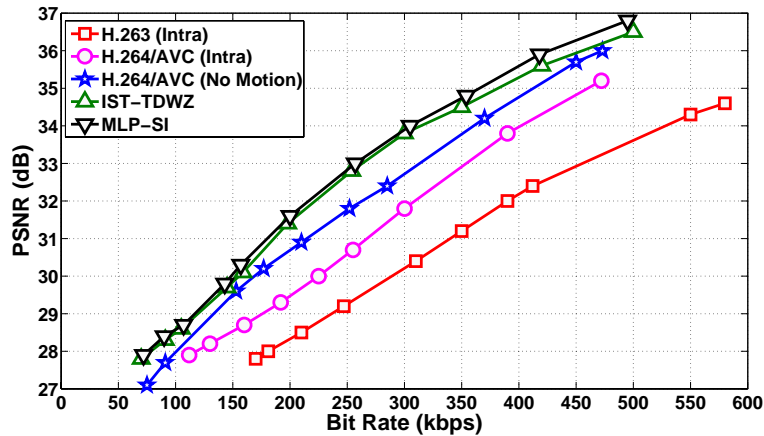
Figure 4.11: Overall rate distortion performance (*Coastguard* at 15 fps)
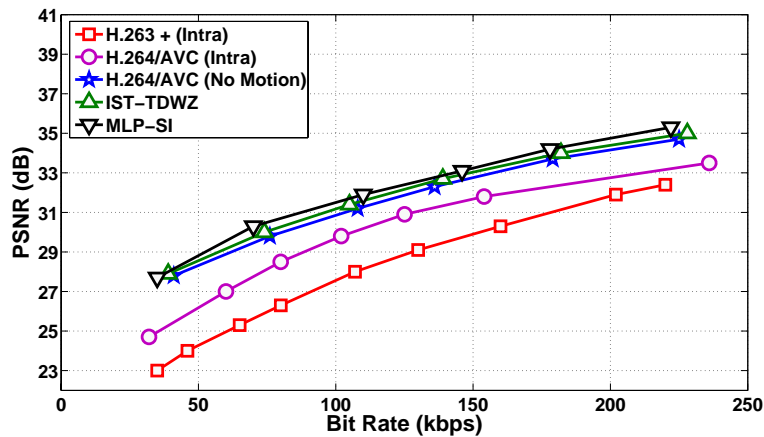
Figure 4.12: Overall rate distortion performance of (*Carphone* at 15 fps)
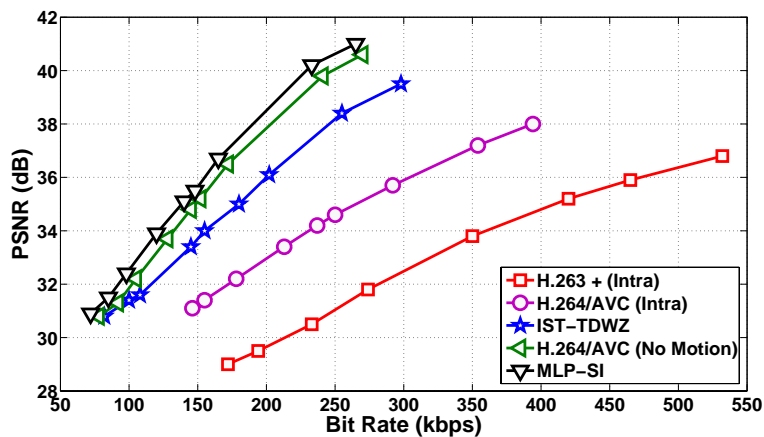
Figure 4.13: Overall rate distortion performance (*Miss America* at 15 fps)
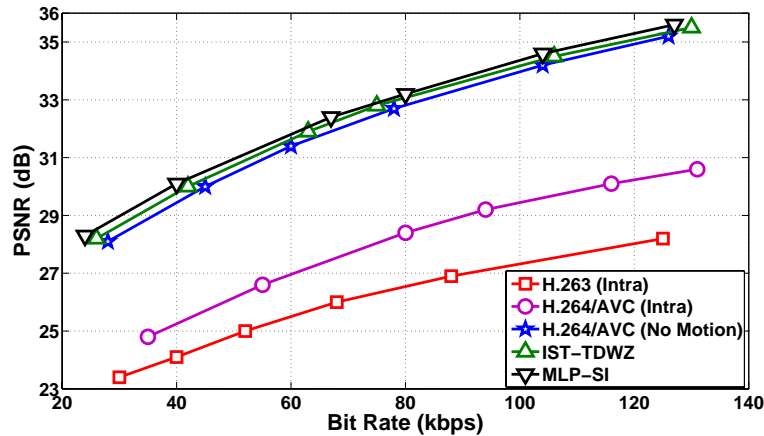
Figure 4.14: Overall rate distortion performance (*Silent* at 15 fps)

(b) At 30 fps, the frames are more closer with less complex motion characteristics, the proposed codec outperforms all codec schemes for the *Foreman, Carphone* and *Silent* sequences. The PSNR improvements of the proposed codec over IST-TDWZ are 0.59 dB, 0.53 dB and 0.572 dB respectively.
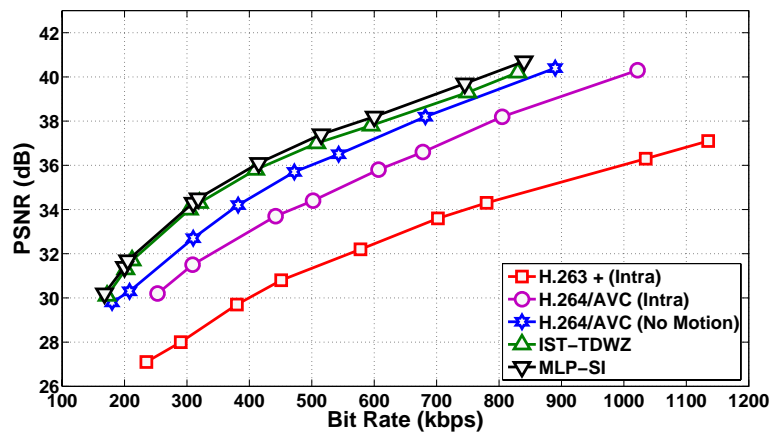
Figure 4.15: Overall rate distortion performance (*Foreman* at 30 fps)

Hence in general, the proposed MLP-SI scheme shows improved RD performance as compared to other schemes in all considered video sequences.

The PSNR (dB) gain of our scheme over IST-TDWZ is given in Table 4.2 with different bitrate and for different sequences. It can be observed that the PSNR gain increases along with bit rate.
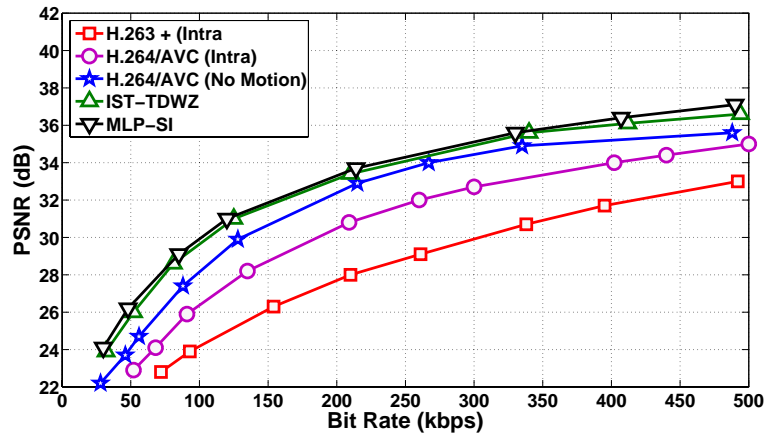
Figure 4.16: Overall rate distortion performance (*Carphone* at 30 fps)
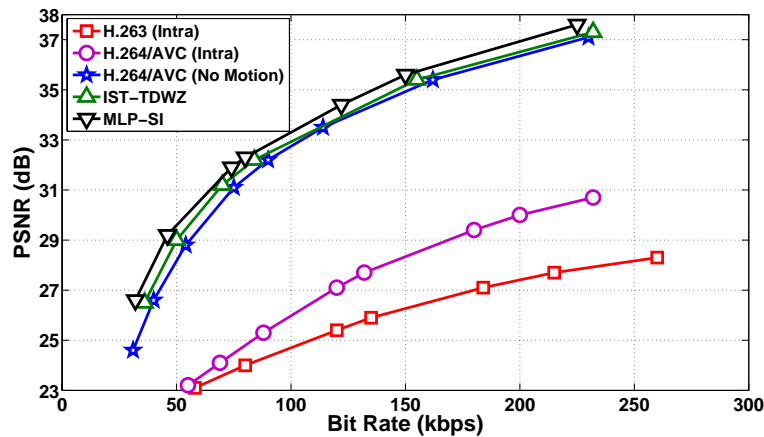
Figure 4.17: Overall rate distortion performance (*Silent* at 30 fps)

Table 4.2: PSNR (dB) gain of MLP-SI over IST-TDWZ at different bit rate in different video sequences

Sequences (15 fps)	Bit Rate				
	100	200	300	400	500
Foreman	0.478	0.513	0.451	0.43	0.418
Carphone	0.346	0.381	0.414	0.452	0.461
Coastguard	0.242	0.281	0.346	0.39	0.412

Experiment 4: Analysis of number of requests per SI frame

This experiment evaluates the number of requests to be made in each frames during the decoding. At the decoder side of the DVC, the parity bits are successively used to correct the errors between WZ frame and SI frame to improve

85

the decoding quality. It is important to study the variation of number requests per SI frame which has a significant impact on decoder complexity. The frame number vs. the number of requests generated in MLP-SI and IST-TDWZ schemes for *Coastguard*, *Foreman*, and *Miss America* sequences are shown in Figures 4.18 - 4.20 respectively. For *Coastguard* sequence, the proposed MLP-SI has a maximum of 834 requests in contrast to 882 in IST-TDWZ for frame number 30. The reason for overshoot in number of requests per SI frame is due to the fact that there is a crossing of two boats in the *Coastguard* sequence. At that point, the camera moves up fast (tilt up) and starts following the bigger boat towards the right direction (pan-right). This fast or strong tilt up occurs from frame number 30-40; as a result due to inconsistent motion the SI generation is not proper in any scheme. So to rectify the SI frame, more number of requests is generated by the schemes to receive more parity bits. Similarly, the maximum requests is 756 in MLP-SI and that in IST-TDWZ is 793 for frame 110 for *Foreman* sequence. For *Miss America* sequence the maximum requests is 823 in contrast to 867 in IST-TDWZ for frame number 116. This comparison reveals the superiority of SI frames generated in MLP-SI as that of IST-TDWZ.
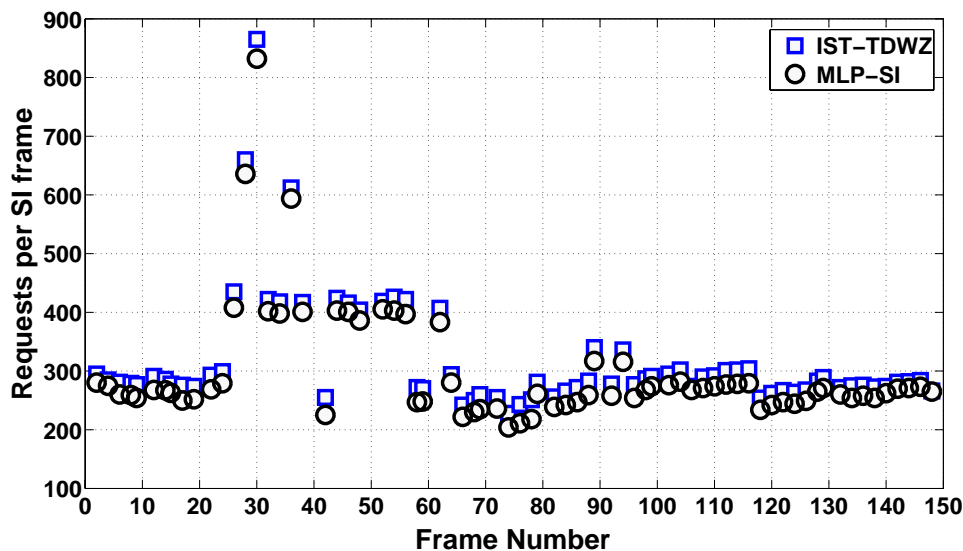
Figure 4.18: Number of requests per SI frame (*Coastguard* at 15 fps)
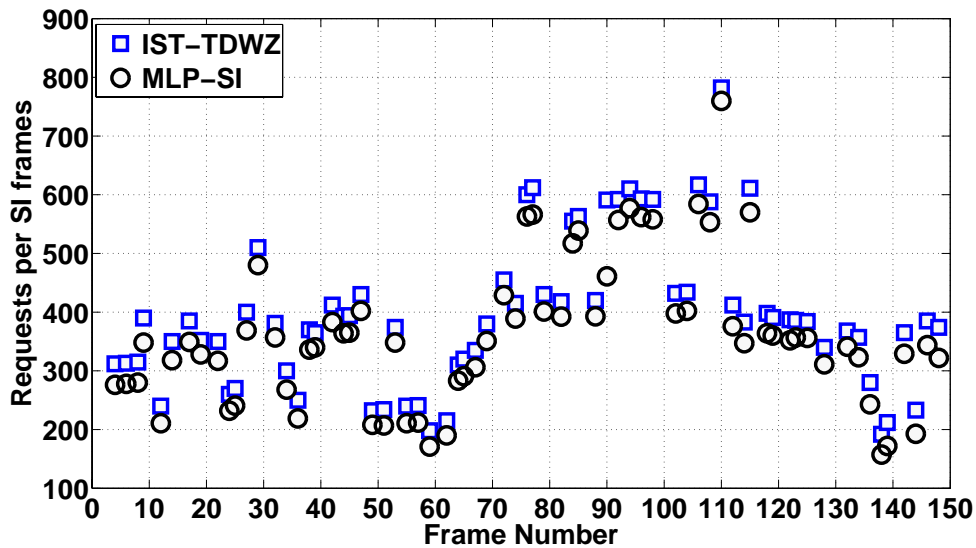
Figure 4.19: Number of requests per SI frame (*Foreman* at 15 fps)
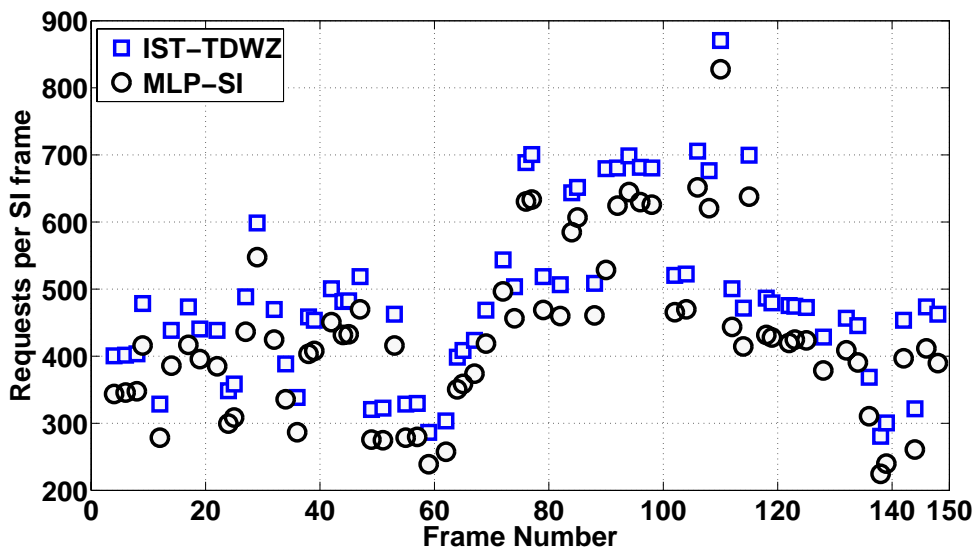
Figure 4.20: Number of requests per SI frame (*Miss America* at 15 fps)

Experiment 5: Temporal evaluation of PSNR (dB)

The SI generation in both the schemes are improved using the parity bits. The final corrected frame will be of higher quality if its PSNR is higher. To study the temporal frame quality, comparative analysis of PSNR (dB) is given in Figures 4.21 and 4.22 for *Foreman* and *Coastguard* sequences respectively.

It may be observed that MLP-SI outperforms IST-TDWZ with respect to PSNR (dB) in both the cases. For computing the average PSNR for all the
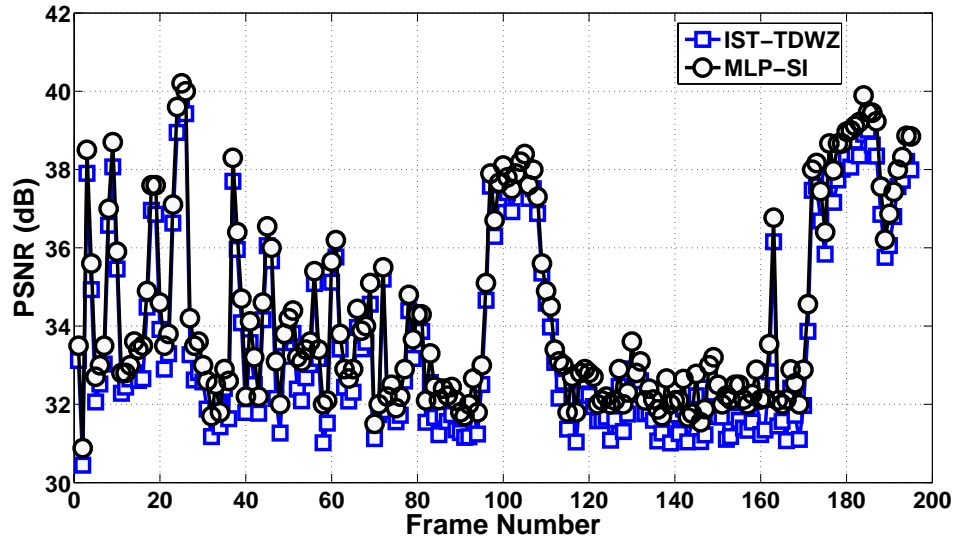
Figure 4.21: Temporal evaluation (*Foreman* at 30 fps)
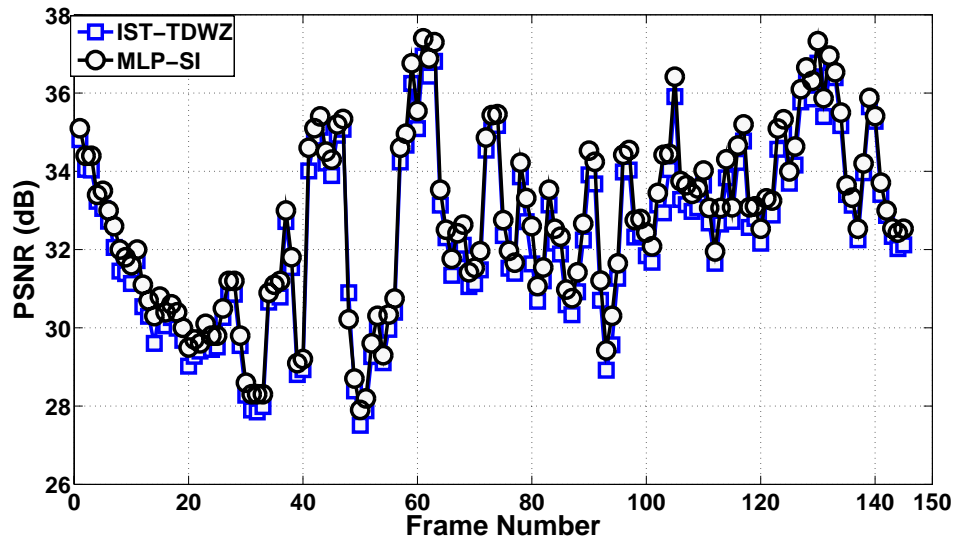
Figure 4.22: Temporal evaluation (*Coastguard* at 30 fps)

decoded frames both key frames and WZ frames are taken into consideration and is shown in Table 4.3. The proposed codec has gain of 0.46 dB, 0.38 dB and 0.41 dB over IST-TDWZ for *Foreman*, *Coastguard* and *Miss America* sequences respectively.

Table 4.3: PSNR comparison of MLP-SI for decoded frames

Q_i	Foreman		Coastguard		Miss America	
	IST-TDWZ	MLP-SI	IST-TDWZ	MLP-SI	IST-TDWZ	MLP-SI
1	27.94	28.23	25.81	26.13	26.23	26.51
2	28.92	29.34	26.42	26.73	27.42	27.63
3	29.26	29.48	28.33	28.54	28.31	28.67
4	32.04	32.38	29.28	29.57	29.44	29.73
5	32.13	32.41	30.72	31.02	29.72	30.01
6	33.38	33.62	32.24	32.63	32.61	32.92
7	35.56	35.84	34.21	34.68	34.74	35.23
8	37.42	37.88	36.27	36.65	36.71	37.05

Experiment 6: Analysis of decoding time requirement

Normally, in a WZ video coding scenario the decoder is much more complex than the encoder, due to the decoding task of both key frames and WZ frames. In order to compare the decoder performance, the total decoding time of complete video sequences in different codecs are measured using machine with Intel i3 processor (3.4 GHz) with 4GB RAM. The complete time comparison is shown in Table 4.4. The quantization matrices used are Q_4 and Q_8 defined in [91]. It may be observed that the proposed MLP-SI shows improvement in decoding time across video sequences and quantization used.

Table 4.4: Decoding time comparison of MLP-SI

Sequences	Total decoding time(in seconds)	
	IST-TDWZ	MLP-SI
Foreman (Q_4)	1124	1088
Miss America (Q_4)	422	337
Coastguard (Q_4)	717	621
Foreman (Q_8)	3698	3565
Miss America (Q_8)	1207	1127
Coastguard (Q_8)	3013	2953

4.4 Summary

This chapter presents an improved SI generation scheme in a DVC framework. The scheme utilizes a MLP to predict the WZ frames. It accepts 8×8 blocks from predecessor and successor key frames of a WZ frame and estimates the

corresponding 8×8 block of WZ frame. When all blocks are estimated, the WZ frame is ensembled. The MLP is trained offline using training patterns selected from frames with numerous motion patterns. Further, these frames are considered from different video sequences. Subsequently, the trained MLP is used for SI generation. The proposed scheme is validated with respect to training convergence, RD performance, PSNR performance, number of requests, decoding time complexity etc. Comparative analysis has been performed with standard video codecs using standard video sequences. The proposed MLP-SI scheme is shown to have superior performance over its competitive schemes with respect to both qualitative and quantitative measures.

Chapter 5

MSVR based side information generation with adaptive parameter optimization

This chapter presents an improved side information (SI) generation scheme using multivariable support vector regression (MSVR) in distributed video coding (DVC). The multivariable support vector regression (MSVR) method suitably formulated so that it is applicable to our problem. Furthermore, the regularization parameter C and kernel parameter σ of MSVR are optimized through particle swarm optimization (PSO) technique using mean square error (MSE) as the fitness function. For generating SI in DVC, MSVR model takes the non-overlapping 8×8 blocks of two decoded key frames as inputs and predicts the 8×8 block of Wyner-Ziv (WZ) frame. The training for MSVR is made offline with known training patterns. The proposed scheme shows superior performance in terms of both coding performance and improved perceptual quality as compared to its competitive schemes.

The rest of the chapter is organized as follows. Section 5.1 presents the proposed side information generation using MSVR. The proposed parameter optimization in MSVR using PSO is presented in Section 5.2. Simulation results and comparative analysis of the suggested scheme is discussed in Section 5.3. Finally, Section 5.4 summarizes the chapter.

5.1 Proposed MSVR based side information generation

The regression problem in SVR solves a single-input single-output (SISO) problem. In our case, the input and observable output both are vectors. So a variant of SVR to solve multi variable problem, MSVR is utilized in our case.

In the proposed scheme, to generate side information a 8×8 WZ frame block is predicted from given two 8×8 key frame blocks. To generate the training patterns, 8×8 blocks from key frames and WZ frames are collected from different video sequences with various motion patterns. The 8×8 blocks of two key frames are used as input and corresponding 8×8 WZ frame block is used as the target for training. The advantage of MSVR approach is that it requires less number of features for training to effectively predict the desirable target. The block diagram of the proposed MSVR-SI scheme is shown in Figure 5.1. For completeness, the theoretical background of MSVR followed by the formulation of the proposed Algorithm are discussed as follows.
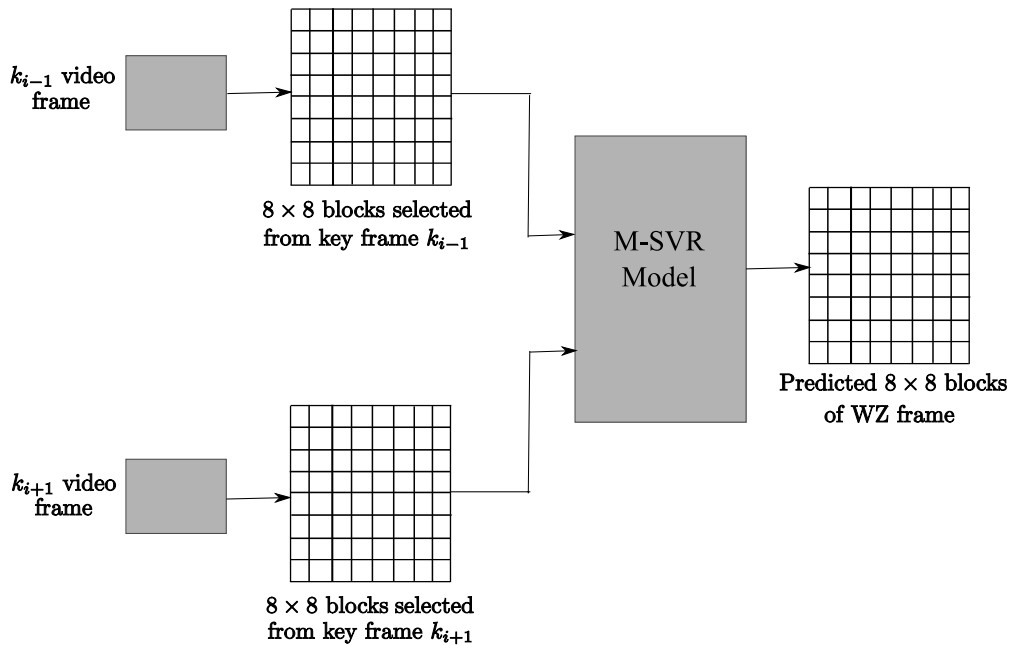
Figure 5.1: Architecture of MSVR-SI scheme

Multivariable Support Vector Regression

Support vector regression (SVR) technique is a theoretically well motivated learning algorithm with the ability to provide high dimensional function estimation [117]. Unidimensional regression problem is regarded as finding the mapping between an incoming vector and an observable output from a given set of samples, $\{(x_i, y_i)\}_{i=0}^{n}$, $x_i \in R^d, y_i \in R$, $i = 0,1,...,n$.

In order to address non-linear regression problem, the input space is mapped to a high dimensional feature space through a nonlinear function $\phi(x)$. The standard SVR problem finds the regressors w and b that minimizes,

$$\frac{1}{2} \, ||w||^2 + C \, \sum_{i=1}^{n} L_\epsilon \, (\, y_i \, - (\phi(x_i)^T \, w + \, b)) \tag{5.1}$$

where, w and b are the high dimensional weight vector and bias respectively. The parameter C controls the trade-off between regularization term and error reduction term. $L_\epsilon(\cdot)$ is known as Vapnik ϵ - insensitive loss function which is defined as,

$$L_\varepsilon(v) \; = \; \begin{cases} 0 \;\; \text{for } |v| \; < \; \varepsilon \\ |v| \; - \; \varepsilon \text{ for } |v| \;\; \geq \; \varepsilon \end{cases} \tag{5.2}$$

The SVR problem can be solved using only inner products of two elements of the feature space, where we need not know the non-linear mapping explicitly. Therefore, we only need to specify a kernel function,

$k : X \times X \rightarrow R$ such that,

$$k(x, z) = \phi \, (x)^T \, \phi(z) \tag{5.3}$$

where, x and z are in input vector space X. In this work, we have used a Gaussian kernel for MSVR defined as,

$$k(x_i, x_j) \; = \; \exp{(-\frac{\|x_i - x_j\|^2}{2\sigma^2})} \tag{5.4}$$

where, σ is the kernel parameter of MSVR.

The concept of MSVR was first introduced by Sanchez et al. [118]. In their first contribution they have proposed a generalization of SVR to solve the problem of regression for multiple variables. Later they have contributed SVM multi regression for non-linear channel estimation in MIMO system [119].

In our case also the observable output is a vector denoted as $y \in R^Q$. So the multidimensional regression problem finds the regressors w^j and b^j for j^{th} component of output vector y $(j=1,2,....Q)$ that minimizes,

$$L_p(W,b) = \frac{1}{2} \sum_{j=1}^{Q} ||w^j||^2 + C \sum_{i=1}^{n} L(u_i) \tag{5.5}$$

where, the Vapnik ϵ- insensitive loss function L extended to multiple dimension is defined as follows,

$$L(u_i) = \begin{cases} 0 & \text{if } u_i < \epsilon \\ u_i^2 - 2u_i\epsilon + \epsilon^2 & \text{if } u_i \geq \epsilon \end{cases} \tag{5.6}$$

In the above expression, $u_i = ||e_i|| = \sqrt{e_i^T e_i}$, $e_i^T = y_i^T - \Phi(x_i)^T W - b^T$, $W = [w^1, ..., w^Q]$, $b = [b^1, ..., b^Q]^T$

Some of the optimization problems are solved using iterative procedures that rely in each iteration on the previous solution. This type of optimization is often referred to as iterative weighted least square procedure (IRWLS). This IRWLS procedure is adopted in our work to find the regressors W and b. Now, we modified (5.5) using a first order Taylor expansion as,

$$\begin{aligned} L_P'(W,b) &= \frac{1}{2} \sum_{j=1}^{Q} ||w^j||^2 + C \sum_{i=1}^{n} [L(u_i^{(m)} + u_i - u_i^{(m)})] \\ &= \frac{1}{2} \sum_{j=1}^{Q} ||w^j||^2 + C \sum_{i=1}^{n} [L(u_i^{(m)}) + \frac{dL(u_i)}{du_i}\bigg|_{u_i^m} (u_i - u_i^{(m)})] \end{aligned} \tag{5.7}$$

where, $u_i^{(m)} = \left\|e_i^{(m)}\right\| = \sqrt{(e_i^{(m)})^T e_i^{(m)}}$, $(e_i^m)^T = y_i^T - \phi(x_i)^T W^m - (b^m)^T$.

The term $u_i - u_i^{(m)}$ in (5.7) is not in quadratic form, so to make it quadratic it can be formulated as,

$$\begin{aligned} (u_i - u_i^{(m)}) &= \frac{(u_i - u_i^{(m)})(u_i + u_i^{(m)})}{u_i + u_i^{(m)}} \\ &\approx \frac{u_i^2 - u_i^{(m)\,2}}{2u_i^m} \end{aligned} \tag{5.8}$$

where, u_i in the denominator is approximated as u_i^m. Using a quadratic approximation $\frac{u_i^2 - u_i^{(m)\,2}}{2u_i^m}$ of $(u_i - u_i^{(m)})$, (5.7) can be written as,

$$L_P''(W, b) = \frac{1}{2} \sum_{j=1}^{Q} ||w^j||^2 + C \sum_{i=1}^{n} \left[L(u_i^{(m)}) + \left. \frac{dL(u_i)}{du_i} \right|_{u_i^{(m)}} \frac{u_i^2 - u_i^{(m)\,2}}{2u_i^{(m)}} \right]$$

$$= \frac{1}{2} \sum_{j=1}^{Q} ||w^j||^2 + C \sum_{i=1}^{n} L(u_i^{(m)}) + C \sum_{i=1}^{n} \left. \frac{dL(u_i)}{du_i} \right|_{u_i^{(m)}} \frac{u_i^2}{2u_i^{(m)}}$$

$$- C \sum_{i=1}^{n} \left. \frac{dL(u_i)}{du_i} \right|_{u_i^{(m)}} \frac{u_i^{(m)\,2}}{2u_i^{(m)}}$$

$$= \frac{1}{2} \sum_{j=1}^{Q} ||w^j||^2 + \frac{1}{2} \sum_{i=1}^{n} a_i u_i^2 + CT \tag{5.9}$$

where,

$$a_i = \frac{C}{u_i^m} \left. \frac{dL(u_i)}{du_i} \right|_{u_i^m} = \begin{cases} 0, & u_i^m < \varepsilon \\ \frac{2C(u_i^m - \varepsilon)}{u_i^m} & u_i^m \geq \varepsilon \end{cases} \tag{5.10}$$

and CT is a sum of constant terms that do not depend on either W or b. Thus the minimization of $L_P(W, b)$ reduces to minimization of $L_P''(W, b)$. To minimize $L_P''(W, b)$, we first compute the gradient of $L_P''(W, b)$ with respect to w^j and b^j.

$$\nabla_{w^j} L_P'' = w_j - \sum_i \phi(x_i) a_i (y_{ij} - \phi(x_i)^T w^j - b^j) = 0 \tag{5.11}$$

$$\nabla_{b^j} L_P'' = - \sum_i a_i (y_{ij} - \phi(x_i)^T w^j - b^j) = 0 \tag{5.12}$$

Combining the expressions in (5.11) and (5.12) we obtain a linear system represented as,

$$\begin{bmatrix} K + D_a^{-1} & 1 \\ a^T K & 1^T a \end{bmatrix} \begin{bmatrix} \beta^j \\ b^j \end{bmatrix} = \begin{bmatrix} y^j \\ a^T y^j \end{bmatrix} \tag{5.13}$$

where, $K = (k_{ij}) = (k(x_i, x_j))$ is known as the kernel matrix and $\beta = (\beta^1, \beta^2, ...\beta^Q)$, $\beta^j = (\beta_1^j \beta_2^j .... \beta_n^j)^T$, where β_i^j is defined by $w^j = \sum_{i=1}^{n} \phi(x_i) \beta_i^j$, $y_j = [y_{1j}, y_{2j}, ..., y_{nj}]$, $a = [a_1, a_2, ..., a_n]^T$, and

$$(D_a)_{ij} = \delta(i, j) = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

For simplicity, the formulation of IRWLS procedure is summarized in *Algorithm* 5.

95

Algorithm 5

1: Choose a kernel function k and compute the kernel matrix K which can be represented as : $K = (k_{ij}) = (k(x_i, x_j)), x_i, x_j \in X$

2: Set $m = 0$, $\beta^{(m)} = 0$, $b^{(m)} = 0$
where, $\beta = (\beta^1, \beta^2, ..., \beta^Q)$, $b = (b^1, b^2, ..., b^Q)^T$, $\beta^j = (\beta_1^j, \beta_2^j, ..., \beta_n^j)^T$ and β_i^j is defined as,

$$w^j = \sum_{i=1}^n \phi(x_i) \beta_i^j \qquad (5.14)$$

3: Compute $u_i^{(m)}$ as,

$$u_i^{(m)} = \sqrt{(y_i^{(m)} - b^{(m)})^T (y_i^{(m)} - b^{(m)}) - 2\alpha_i + A_i A_i^T} \qquad (5.15)$$

where, $A_i = \phi^T(x_i)W$ which can be further reduced as,
$A = \sum_{l=1}^n k(x_i, x_l) \beta_l^T$ where, $\beta_l = l^{th}$ row of β and $\alpha = A_i(y_i - b)$

4: Compute a_i using equation 5.10

5: Compute the solution of linear system in equation 5.13 and label as $\beta^{(s)} and b^{(s)}$

6: Define $P^{(m)}$ as,

$$P^{(m)} = \begin{bmatrix} \beta^{(s)} - \beta^{(m)} \\ (b^{(s)} - b^{(m)})^T \end{bmatrix} \qquad (5.16)$$

Compute next approximation to β and b as

$$\begin{bmatrix} \beta^{(m+1)} \\ b^{(m+1)T} \end{bmatrix} = \begin{bmatrix} \beta^{(m)} \\ (b^m)^T \end{bmatrix} + \eta^{(m)} P^{(m)} \qquad (5.17)$$

where, step size $\eta^{(m)}$ is computed using back tracking algorithm given below: if $L_p(\beta^{(m+1)}, b^{(m+1)} < L_p(\beta^{(m)}, b^{(m)})$ then accept the value of η otherwise, multiply η_m by a positive constant less than one and repeat the procedure until a decrease is achieved in the minimizing functional L_p. Here, $L_P(\beta^{(m)}, b^{(m)})$ is defined as,

$$L_P(\beta^{(m)}, b^{(m)}) = \frac{1}{2} \sum_{j=1}^Q \sum_{i=1}^n \sum_{l=1}^n k_{il} \beta_i^{j\,(m)} \beta_l^{j\,(m)} + C \sum_{i=1}^n L(u_i^{(m)}) \qquad (5.18)$$

7: Set $m=m+1$ and repeat steps 3 to 6 until convergence occurs i.e. $\max_j \left\| (\beta^{j\,(m+1)} - \beta^{j\,(m)} \right\| + \left\| b^{(m+1)} - b^{(m)}) \right\| < 0.01$

8: Finally, Compute

$$y^j = \sum_{i=1}^n k(x_i, x_j) \beta_i^j + b^j, j = 1, 2, .., Q \qquad (5.19)$$

where, y^j is the jth output

5.2 Parameter optimization using PSO in MSVR

For accurate prediction results and enhance generalization, the value of parameters in MSVR should be in a specific range. Considering the number of parameters in SVR Vapnik [117] addressed the two relevant parameters namely hyper penalty parameter C and kernel parameter σ. The parameter C controls the trade-off between regularization and error reduction term. When C is small, the error reduction is small and will not get smaller fitting. On the other hand, when C is large, error will be large which will lead to over learning and over generalization. Similarly, the parameter σ also plays a major role in the performance of MSVR. If it is overestimated, the exponential will behave almost linearly and fails to transform to a higher dimensional plane. Similarly, if it is underestimated regularization will be affected and the decision boundary will be sensitive towards noise in training data. If MSVR is applied for SI generation without parameter optimization, the parameters are adjusted heuristically for good result. Such parameters do not provide satisfactory results across video frames with different motion characteristics. So it is necessary to optimize the parameters adaptively to increase the overall efficacy of MSVR. Any global optimization technique like GA [120, 121], PSO [122–124] can be employed for this purpose. We utilize PSO in our case for optimization of C and σ. The purpose of using PSO is that, it has fewer number of adjusting parameters in comparison to GA. Further, PSO shows faster convergence and easier implementation.

In proposed parameter optimization using PSO mean square error (MSE) has been utilized as the objective or fitness function. However, the minimization of MSE function plays an important role in the correct estimation of the regularization parameter C and kernel parameter σ. It has been observed that traditional minimization algorithms fail to achieve the global minimum because MSE function is multimodal in nature. Therefore, any traditional minimization technique using MSE is capable of optimizing either C or σ but not both together. In order to remove this bottleneck, a PSO based global optimization scheme has been proposed to optimized C and σ for minimized MSE. The detailed

optimization methodology is presented in results and discussion.

The MSVR is trained with a random initial set of parameters. The parameters of the MSVR i.e. C and σ are updated till the MSE between the original WZ frame and the predicted WZ frame is below a threshold level, T=0.01. The trained MSVR with the optimized set of parameters is used to predict 8×8 block of WZ frames from two decoded 8×8 blocks of key frames. The flow chart of the proposed parameter optimization using PSO is shown in Figure 5.2.
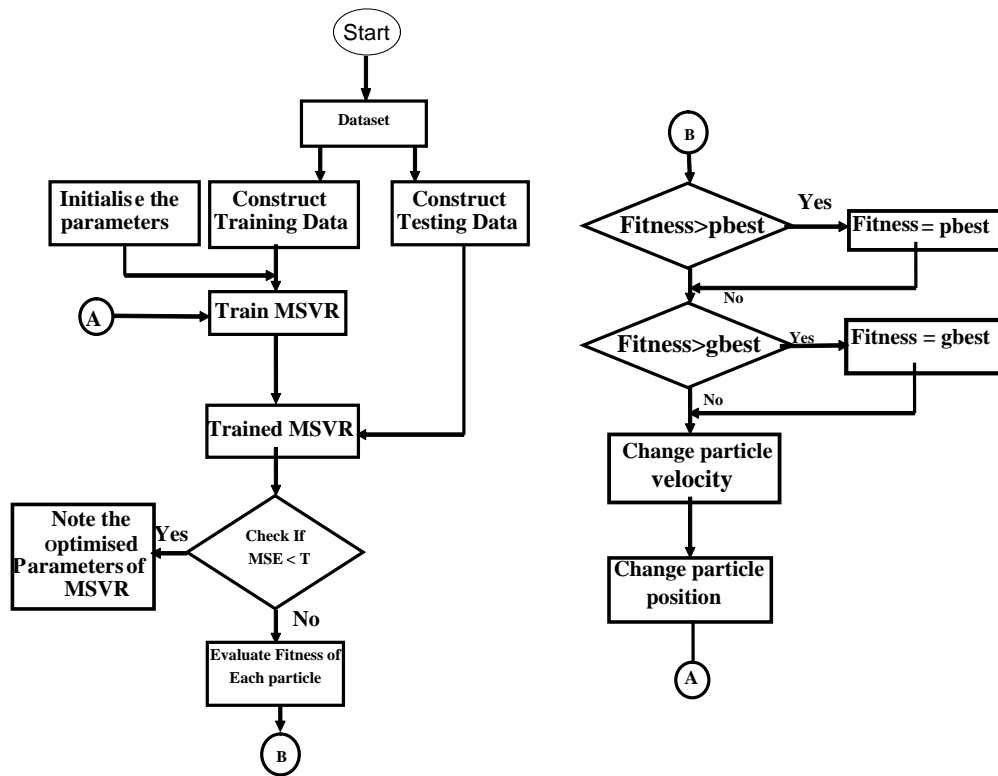
Figure 5.2: Flow chart of parameter optimization using PSO

5.3 Results and discussions

This Section evaluates the performance of the MSVR-SI scheme with the most relevant video codecs. The proposed MSVR-SI scheme has been tested in transform domain Stanford DVC architecture. The overall experimental study is divided into six different experiments and are discussed in sequel.

Experiment 1: MSVR training and parameter optimization methodology

To train the MSVR, a training set constituting of 8×8 blocks from key frames and corresponding 8×8 from WZ frames are selected randomly to use them as inputs and targets respectively. The frames are chosen from different standard video sequences namely, *Foreman, Miss America, Suzie, Container, Coastguard* and *Carphone*. The selected frames include different motion characteristics to constitute a representative training set. A training set of size 4300 blocks is chosen and they constitute 27 frames of *Foreman*, 18 frames of *Miss America*, 13 frames of *Suzie*, 22 frames of *Container*, 20 frames of *Coastgaurd*, and 14 frames of *Carphone* video.

Since both C and σ are important parameters in MSVR, we try to optimize them using PSO. Initially, for a fixed value of σ, PSO tries to optimize value of C with respect to fitness function as MSE (dB). Subsequently, for the observed value of C, the σ is optimized by PSO. This process is repeated for a set of iteration till the values stabilizes. Figure 5.3 reflects the MSE characteristics with respect to C values for a fixed value of σ=0.50. Similarly, Figure 5.4 shows the MSE characteristics with respect to σ for an observed value C =2.51 in previous iteration. Finally, we obtain the optimum value of C=2.16 and $\sigma = 0.91$ and these values are used in MSVR for comparative analysis.
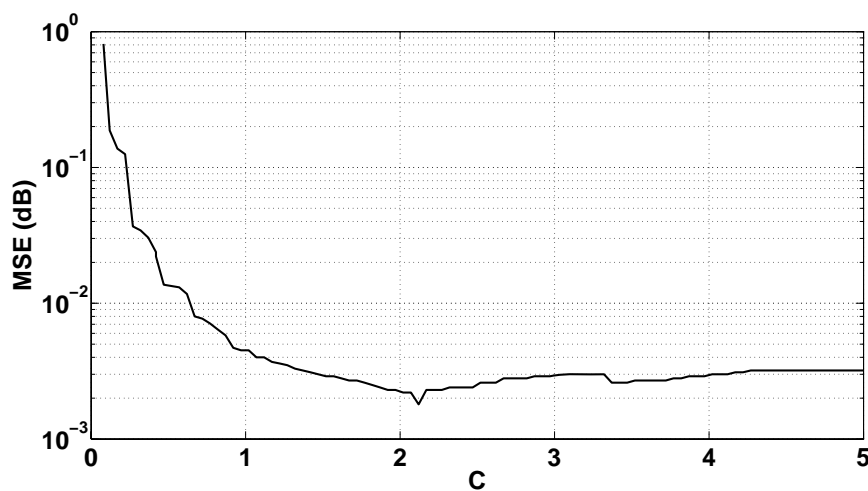
Figure 5.3: C vs. MSE characteristics (for σ =0.50)
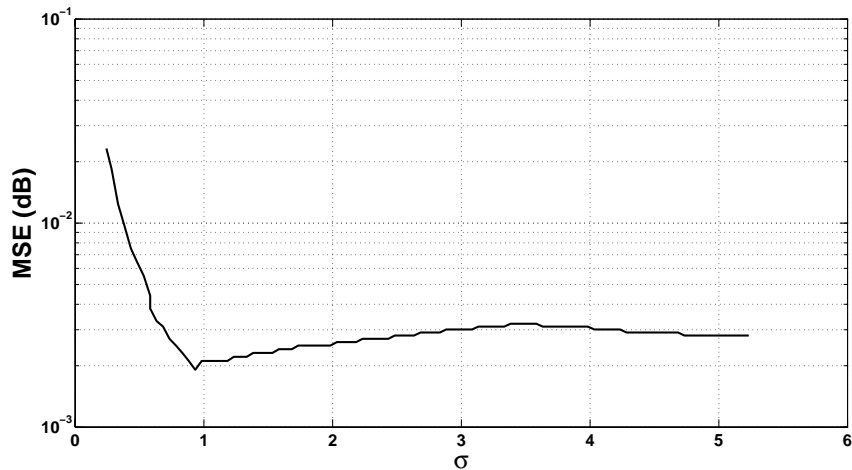
Figure 5.4: σ vs. MSE characteristics (for C =2.51)

Experiment 2: Performance analysis of side information

Among the various competent schemes IST-TDWZ is a recent SI generation scheme in DVC scenario. The MSVR-SI without parameter optimization is also being compared with the proposed scheme. Hence PSNR (dB) for frames *Foreman* and *Coastguard* sequences are compared and shown in Figures 5.5 and 5.6 respectively. It may be observed that the SI frames generated by MSVR-SI
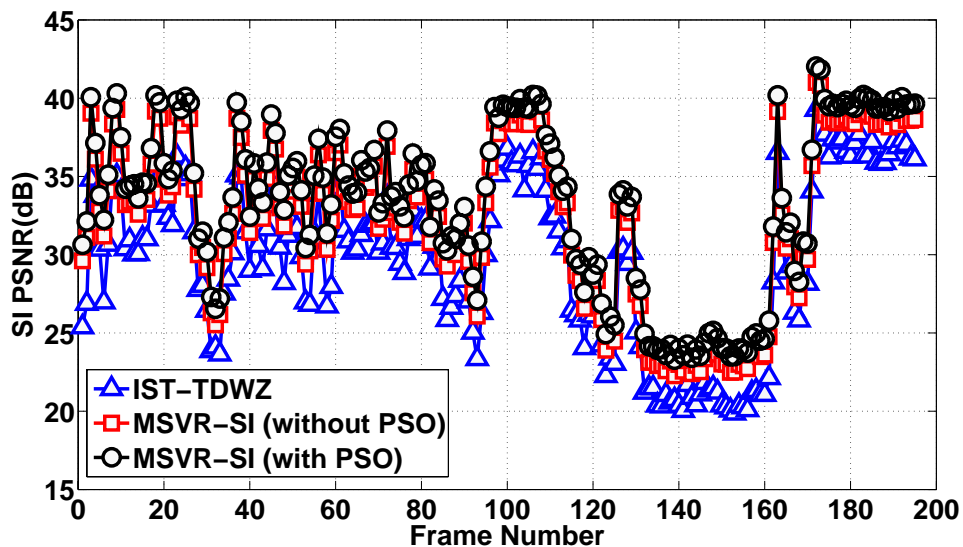
Figure 5.5: PSNR of SI frames (*Foreman* at 15 fps)

scheme has a superior PSNR performance as compared to IST-TDWZ and MSVR-SI scheme without parameter optimization. This reflects the qualitative
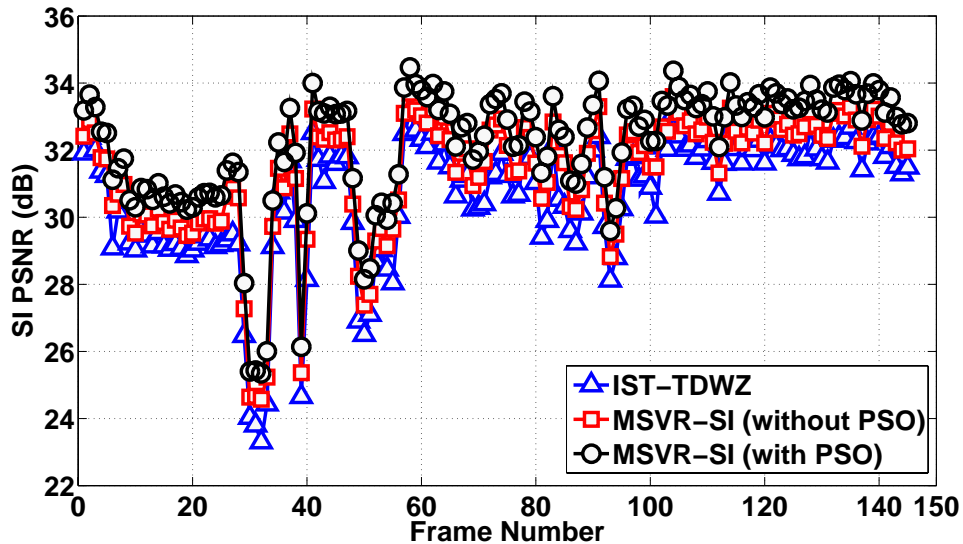
Figure 5.6: PSNR of SI frames (*Coastguard* at 15 fps)

side information of WZ frames. The subjective performance is shown in Figures 5.7 and 5.8 for frame number 84 and 36 of *Foreman* and *Suzie* video respectively. It may be noted that the optimized MSVR-SI frame is visually more appealing as compared to IST-TDWZ and MSVR-SI without parameter optimization.

Experiment 3: Analysis of overall rate distortion performance

To evaluate the overall efficacy of the proposed MSVR-SI generation scheme, it is integrated with the Stanford based DVC architecture. To demonstrate the RD performance of different standard video codecs along with our MSVR-SI scheme, simulation results are shown in Figures 5.9 - 5.16. Following observations have been made;

(a) MSVR-SI has an average PSNR gain of 0.63 dB at lower bit rate and 0.74 dB at higher bit rate as compared to H.264/AVC (no motion) in *Foreman* sequence. At 15 fps, MSVR-SI has a PSNR (dB) gain of 0.514 and 0.63 over IST-TDWZ for *Coastguard* and *Carphone* sequence respectively. At 15 fps, with slow motion characteristics, the proposed MSVR-SI scheme has an average PSNR gain of 0.51 dB over H.264/AVC (no motion) for *Miss America* sequence and 0.573 dB over IST-TDWZ for *Silent* sequence.

(b) At 30 fps, the frames are more closer with less complex motion characteristics

(a) IST-TDWZ   (b) MSVR-SI (without PSO)

(c) MSVR-SI (with PSO)
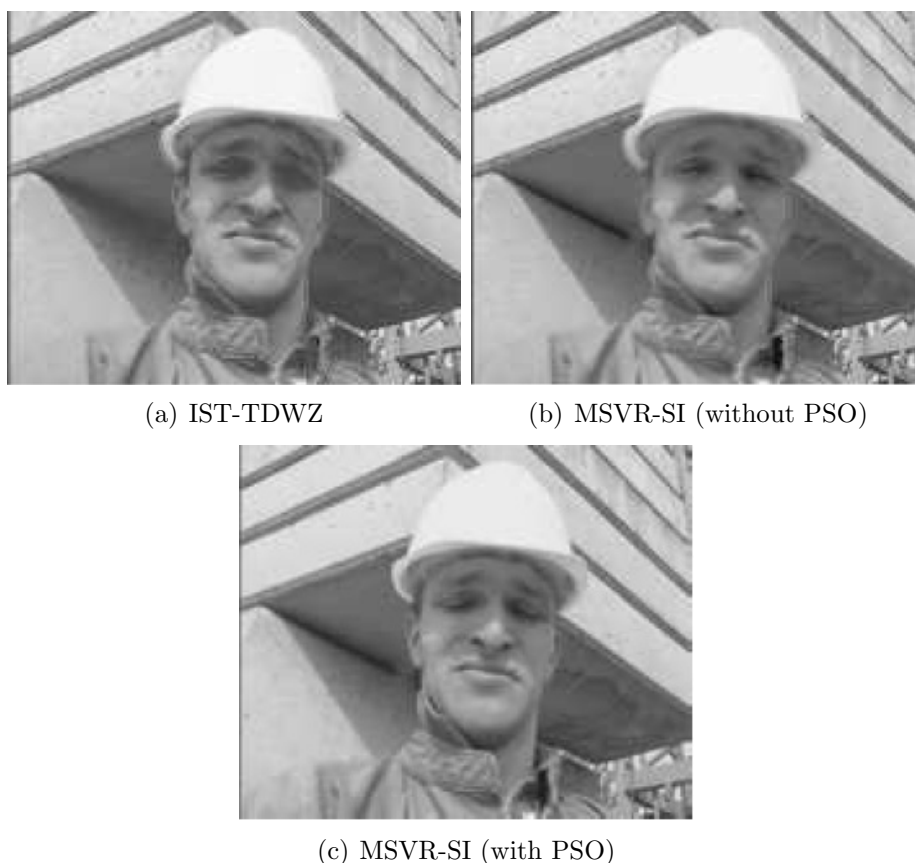
Figure 5.7: Subjective analysis of a decoded SI frame (84th frame of *Foreman*)

the proposed codec outperforms all codec schemes for the *Foreman*, *Carphone* and *Silent* sequences. The PSNR improvements of the proposed codec over IST-TDWZ for that three sequences are 0.79 dB, 0.64 dB and 0.68 dB respectively.

In general, the proposed MSVR-SI scheme shows improved RD performance as compared to other schemes in all video sequences.

Experiment 4: Analysis of number of requests per SI frame

This experiment evaluates the number of requests to be made in each frames during decoding. At the decoder side of the DVC, the parity bits are successively used to correct the errors between WZ frame and SI frame to improve the decoding quality. It is important to study the variation of number requests per SI frame which has a significant impact on decoder complexity. The frame number vs. the number of requests per SI frame generated in MSVR-SI and IST-TDWZ schemes for *Coastguard* and *Foreman* sequences are shown in Figures 5.17 and 5.18

(a) IST-TDWZ   (b) MSVR-SI (without PSO)

(c) MSVR-SI (with PSO)

Figure 5.8: Subjective analysis of a decoded SI frame (36th frame of *Suzie*)
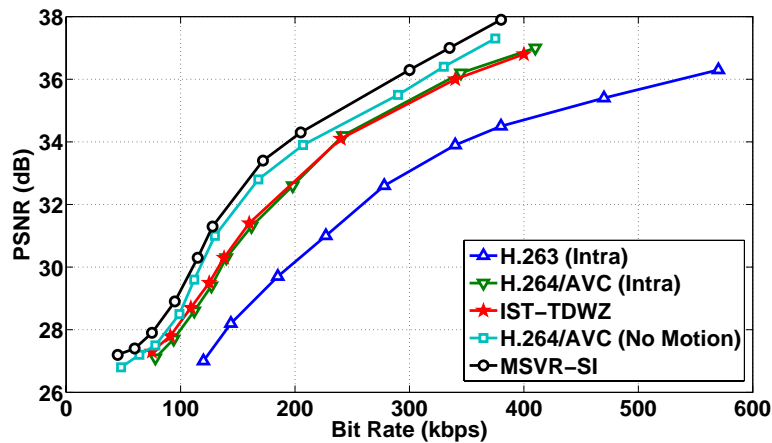
Figure 5.9: Overall rate distortion performance (*Foreman* at 15 fps)

respectively. For *Coastguard* sequence, the proposed MSVR-SI has a maximum of 796 requests in contrast to 882 in IST-TDWZ for frame number 26. Similarly, the maximum requests is 715 in MSVR-SI and that in IST-TDWZ is 793 for frame number 114 for *Foreman* sequence. The reason for overshoot at frames 110-117 is
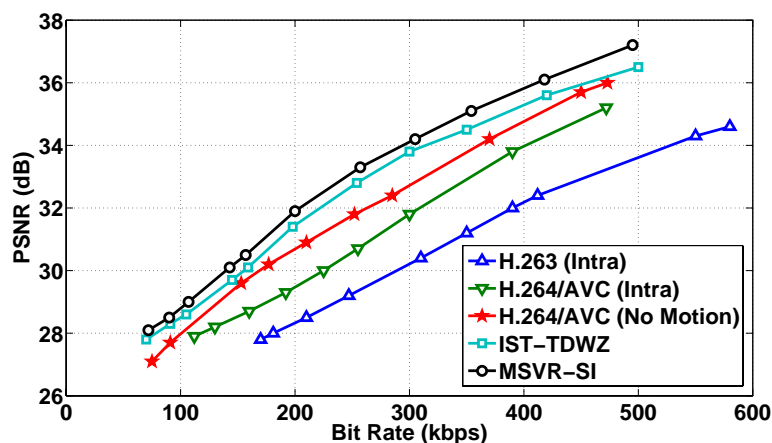
Figure 5.10: Overall rate distortion performance (*Coastguard* at 15 fps)
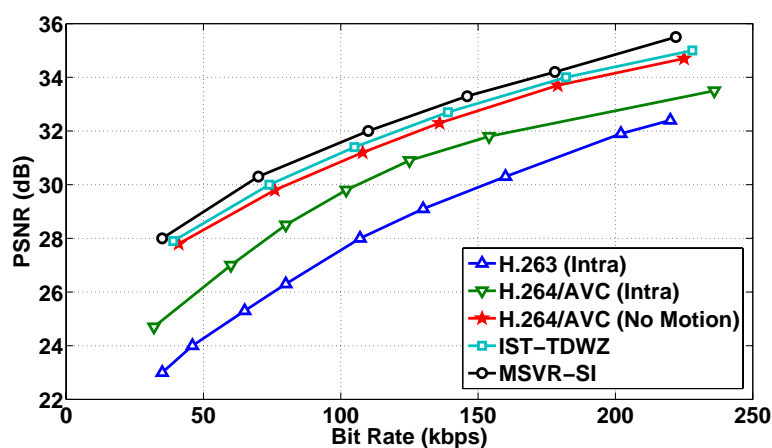
Figure 5.11: Overall rate distortion performance (*Carphone* at 15 fps)
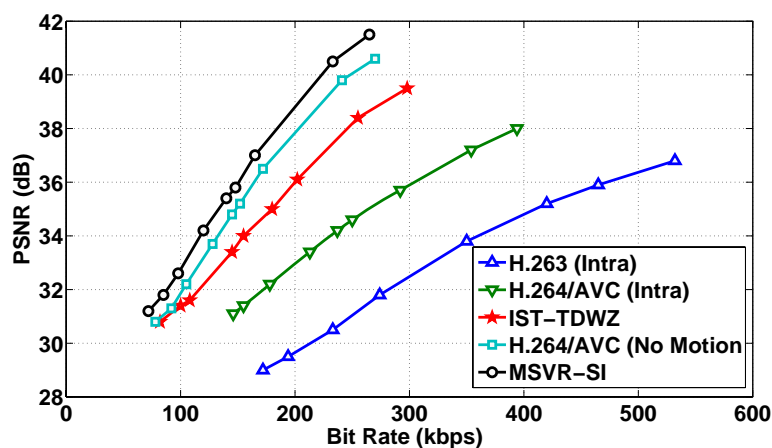
Figure 5.12: Overall rate distortion performance (*Miss America* at 15 fps)
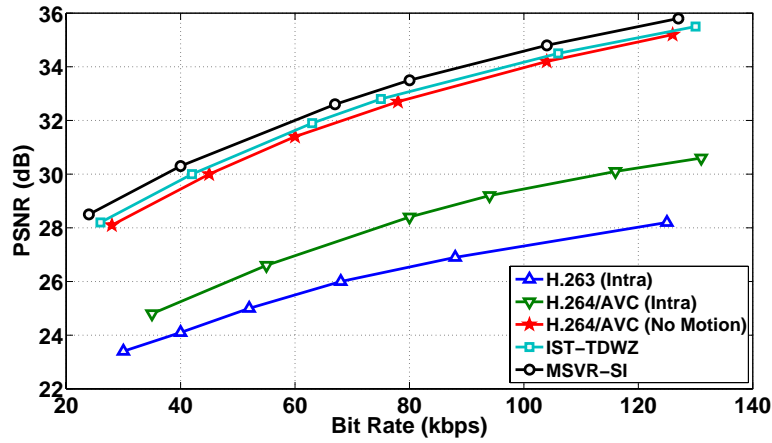
Figure 5.13: Overall rate distortion performance (*Silent* at 15 fps)
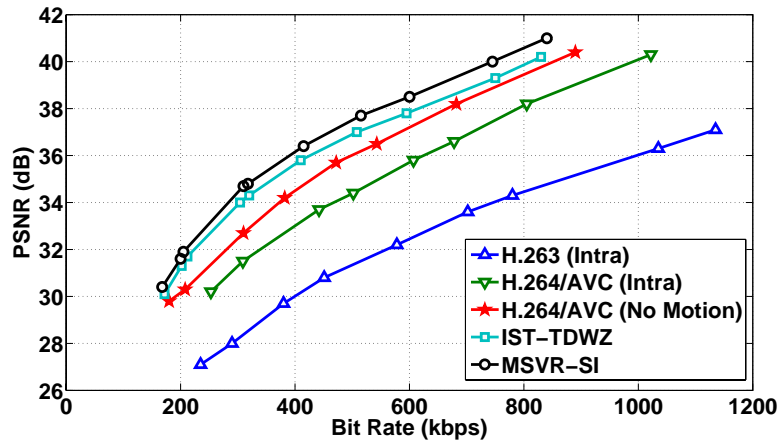
Figure 5.14: Overall rate distortion performance (*Foreman* at 30 fps)
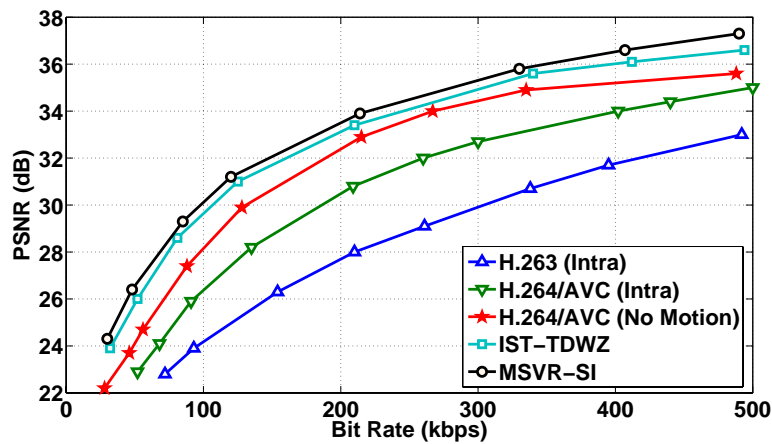
Figure 5.15: Overall rate distortion performance (*Carphone* at 30 fps)
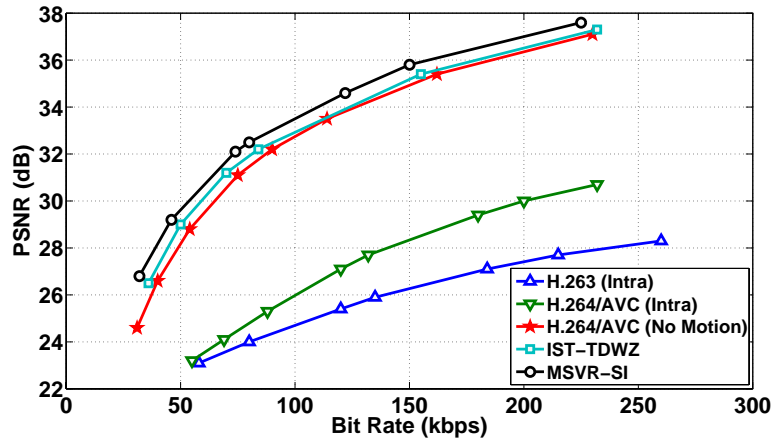
Figure 5.16: Overall rate distortion performance (*Silent* at 30 fps)

due the fact that in frame 110-117 there is a fast change with horizontal panning to scene with a "building under construction". During this there is a fast camera panning, while towards the end of the sequence the camera is still and thus there is almost no motion in the scene and in the camera.
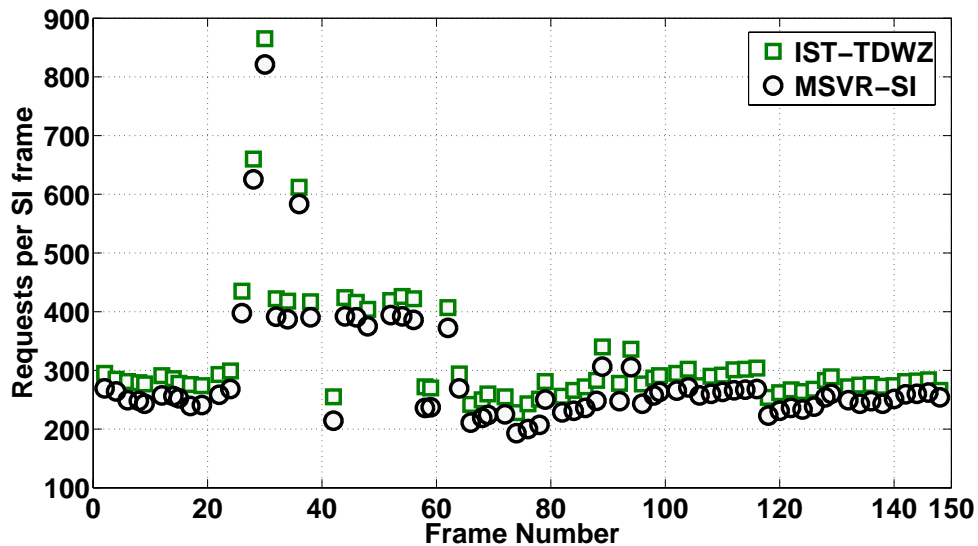
Figure 5.17: Number of requests per SI frame (*Coastguard* at 15 fps)

Experiment 5: Temporal evaluation of PSNR (dB)

To study the temporal frame quality, comparative analysis of over all PSNR (dB) is given in Figures 5.19 and 5.20 for *Foreman* and *Coastguard* sequences respectively.

It may be observed that MSVR-SI outperforms IST-TDWZ with respect to

106
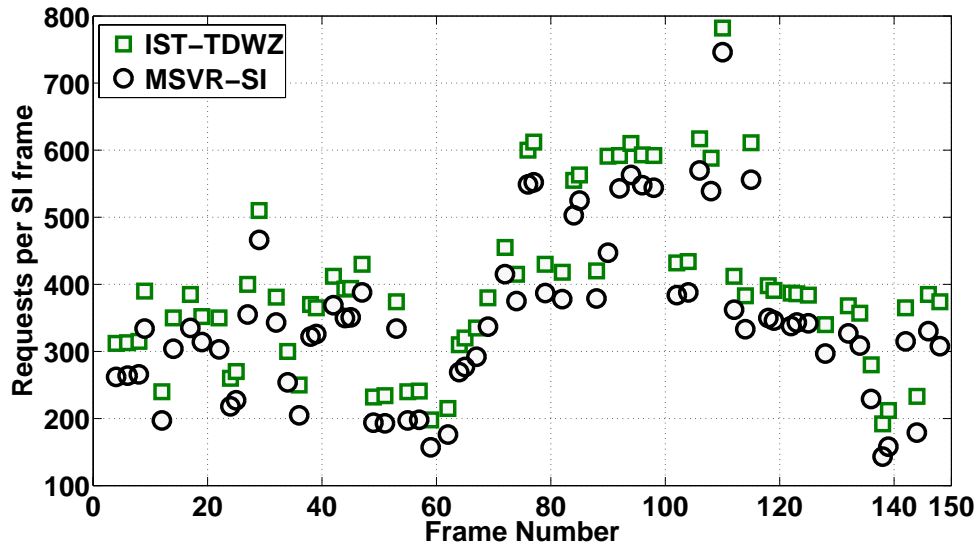
Figure 5.18: Number of requests per SI frame (*Foreman* at 15 fps)

PSNR (dB) in both the cases. The PSNR (dB) gain of MSVR-SI scheme over IST-TDWZ is given in Table 5.1 with different bitrate and for different sequences. It can be observed that the PSNR gain increases along with bit rate.

Table 5.1: PSNR (dB) gain of MSVR-SI over IST-TDWZ at different bit rate in different video sequences

Sequences (15 fps)	Bit Rate				
	100	200	300	400	500
Foreman	0.478	0.523	0.457	0.462	0.44
Carphone	0.42	0.472	0.48	0.53	0.54
Coastguard	0.252	0.281	0.43	0.462	0.48.3

For computing the average PSNR for all the decoded frames both key frames and WZ frames are taken into consideration and is shown in Table 5.2. The proposed codec has gain of 0.61 dB, 0.52 dB and 0.48 dB over IST-TDWZ for *Foreman*, *Coastguard* and *Carphone* sequences respectively.

Experiment 6: Analysis of decoding time requirement

The total decoding time of each video sequence is recorded and listed in Table 5.3. The quantization matrices Q_4 and Q_8 [91] are used in each case. It may be observed that MSVR-SI scheme has an edge over IST-TDWZ in decoding time.
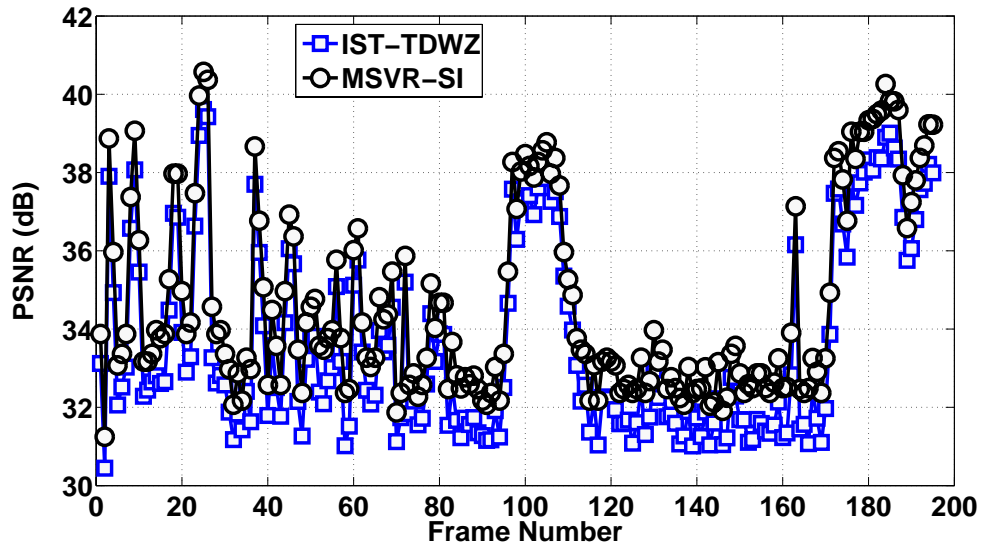
Figure 5.19: Temporal evaluation (*Foreman* at 30 fps)

Table 5.2: PSNR comparison of MSVR-SI for decoded frames

Q_i	Foreman		Coastguard		Carphone	
	IST-TDWZ	MSVR-SI	IST-TDWZ	MSVR-SI	IST-TDWZ	MSVR-SI
1	27.94	28.53	25.81	26.33	24.23	24.81
2	28.92	29.64	26.42	26.94	25.42	25.81
3	29.26	29.73	28.33	28.56	26.31	27.11
4	32.04	32.54	29.28	29.83	27.44	27.92
5	32.13	32.67	30.72	31.34	28.72	29.32
6	33.38	33.81	32.24	32.95	30.61	31.31
7	35.56	36.12	34.21	34.91	32.74	33.56
8	37.42	38.17	36.27	36.82	34.71	35.43

Table 5.3: Decoding time comparison of MSVR-SI

Sequences	Total decoding time(in seconds)	
	IST-TDWZ	MSVR-SI
Foreman (Q_4)	1124	1053
Miss America (Q_4)	422	306
Coastguard (Q_4)	717	595
Foreman (Q_8)	3698	3512
Miss America (Q_8)	1207	1098
Coastguard (Q_8)	3013	2918

5.4   Summary

This chapter presents, an improved SI generation scheme using MSVR based SI generation in DVC. In our work, we have applied the MSVR algorithm
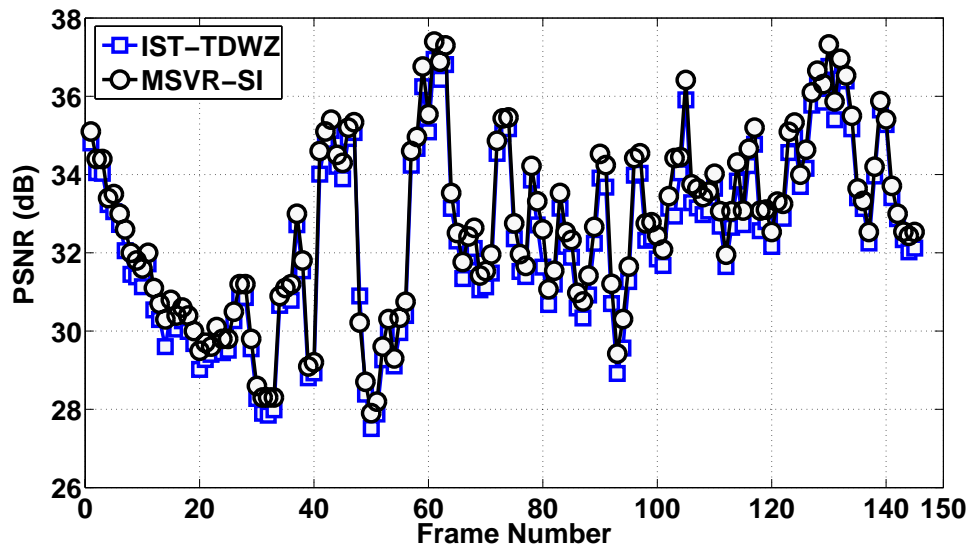
Figure 5.20: Temporal evaluation (*Coastguard* at 30 fps)

and formulate a simplified algorithm for regression problem. Furthermore, the parameters of MSVR are optimized using PSO technique. From different results analysis it has been observed that the proposed MSVR-SI scheme has better standing to its counterpart in terms of both coding efficiency and improved perceptual quality.

Chapter 6

Hybrid schemes formulation out of the suggested schemes

From the distributed video coding (DVC) developments in last few years it has been observed that schemes put more thrust on intra-frame coding and better quality side information (SI) generation. In fact both are interrelated as SI generation is dependent on key frame quality. Hence superior quality key frames generated through intra-key frame coding will in turn are utilized to generate good quality SI frames. As a result, DVC needs less number of parity bits to reconstruct the WZ frames at the decoder. Keeping this mind in this thesis we have proposed two schemes for intra-key frame coding and two schemes for side information generation. The schemes are as follows.

(a) Intra-frame coding using BWT-H.264/AVC (Intra) scheme: Here the Borrows-Wheeler Transform (BWT) is embedded in existing H.264/AVC (Intra) to generate a regularized bit stream prior to compression. This scheme results in higher compression efficiency as well as high quality decoded key frames **(Chapter 2)**.

(b) Intra-frame coding using DBOMP-H.264/AVC (Intra) scheme: The scheme is based on an adaptive dictionary based H.264/AVC (Intra) intra-frame coding. The traditional transform is replaced with a dictionary trained with K-SVD algorithm. The dictionary elements are coded using orthogonal matching pursuit **(Chapter 3)**.

(c) Side information generation using MLP-SI scheme: This scheme utilizes a

multilayer perceptron to estimate SI frames from the decoded key frames block-by-block. The network is trained offline using training patterns from different frames collected from standard video sequences. This is to make the network to adapt to various motion characteristics in real time **(Chapter 4)**.

(d) Side information generation using MSVR-SI scheme: This scheme utilizes an optimized M-SVR to generate SI frames from decoded key frames block-by-block. Like MLP-SI, the training for M-SVR is made offline with known training patterns **(Chapter 5)**.

The suggested schemes are studied in isolation in earlier chapters. The performance comparisons with their competitive schemes have been made with respect to different parameters. The visual performance of these schemes is also made in different frames as well as the sequence as a whole. In this chapter, we combine the proposed intra-frame coding schemes with the side information generation schemes to create four different hybrid schemes in permutation. The hybrid schemes are named as follows.

(a) **BWT-MLP** scheme: The scheme is formulated by combining BWT-H.264/AVC (Intra) intra-frame coding scheme and MLP-SI side information generation scheme.

(b) **BWT-MSVR** scheme: This scheme is a combination of BWT-H.264/AVC (Intra) intra-frame coding scheme and MSVR-SI side information generation scheme.

(c) **DBOMP-MLP** scheme: The scheme is an outcome of putting DBOMP-H.264/AVC (Intra) intra-frame coding followed by MLP-SI side information generation.

(d) **DBOMP-MSVR** scheme: This scheme deals with DBOMP-H.264/AVC (Intra) intra-frame coding and MSVR-SI side information generation together.

The hybrid schemes are incorporated into the Stanford based architecture of DVC and simulation has been carried out on standard video sequences. The performance analysis with respect to overall rate distortion, number requests per SI frame, temporal evaluation, and decoding time requirement has been made to derive an overall conclusion. The performance analysis is described below.

(a) **Analysis of overall rate distortion performance** :

To demonstrate the RD performance of different hybrid schemes, simulation results are shown in Figures 6.1- 6.7. Following observations are derived.

(i) At 15fps, with slow motion characteristics, the hybrid scheme OMP-MSVR has an average PSNR gain of 0.54 dB, 0.9 dB, and 1.73 dB over hybrid schemes BWT-MSVR, DBOMP-MLP, and BWT-MLP for *Miss America* sequence. For *Carphone* sequence the OMP-MSVR has an average PSNR gain of 0.46 dB, 0.7 dB, and 1.13 dB over BWT-MSVR, DBOMP-MLP, and BWT-MLP respectively. The gain for OMP-MSVR is observed to be 0.53 dB, 0.68 dB, and 1.22 dB over BWT-MSVR, DBOMP-MLP, and BWT-MLP.

(ii) At 15 fps for the complex sequence like *Foreman*, the hybrid scheme DBOMP-MSVR has an average PSNR gain of 0.75 dB, 1.38 dB, and 1.64 dB over DBOMP-MLP, BWT-MSVR, and BWT-MLP respectively. However, during this experiment it has been observed that at 15fps, the hybrid scheme BWT-MSVR fails as compared to DBOMP-MLP is due to, in *Foreman* sequence some kind of diagonal motion is observed. So DBOMP-MLP is a good candidate for this as compared to BWT-MSVR. For all other cases, BWT-MSVR has superior performance as compared to DBOMP-MLP.

(iii) At 30 fps, the frames are more closer with less complex motion characteristics, the hybrid codec DBOMP-MSVR outperforms all hybrid codec for *Foreman*, *Carphone* and *Silent* sequences. The PSNR improvements of the DBOMP-MSVR are 0.47 dB, 0.76 dB, and 1.12 dB over DBOMP-MLP, BWT-MSVR, and BWT-MLP for
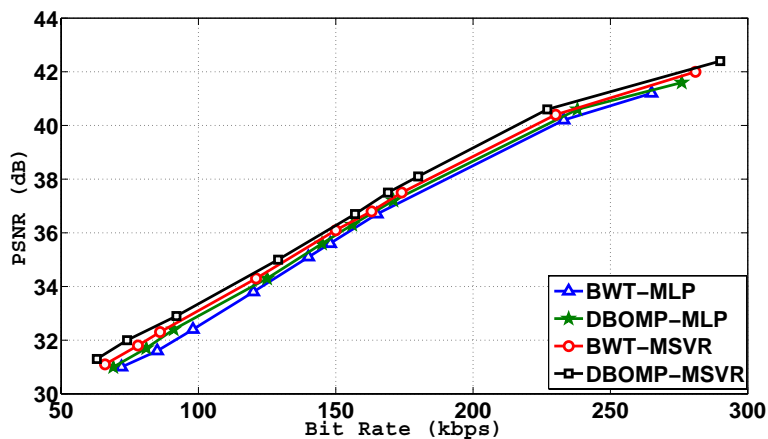
Figure 6.1: Overall rate distortion performance (*Miss America* at 15 fps)
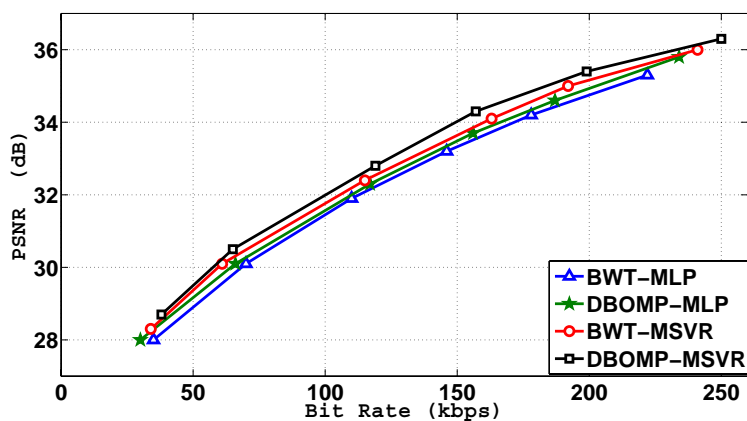
Figure 6.2: Overall rate distortion performance (*Carphone* at 15 fps)
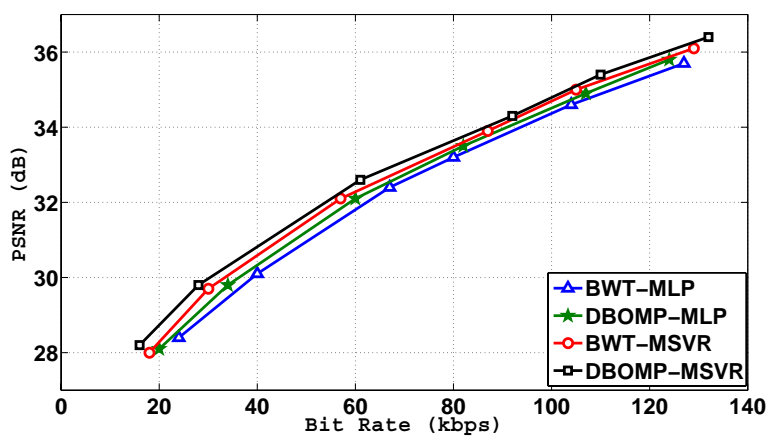
Figure 6.3: Overall rate distortion performance (*Silent* at 15 fps)
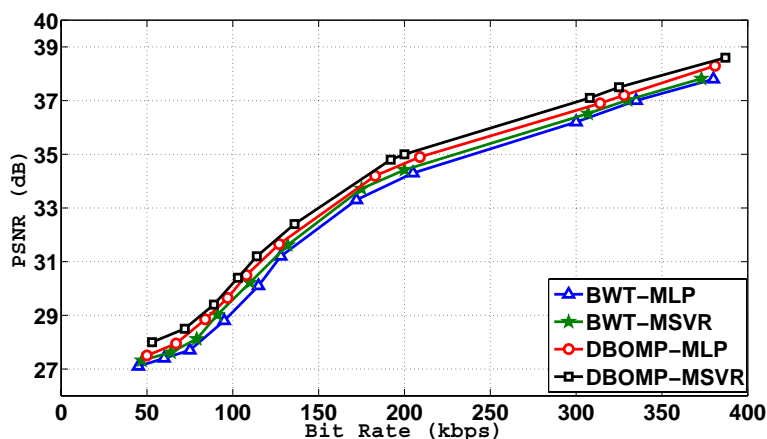
Figure 6.4: Overall rate distortion performance (*Foreman* at 15 fps)

Foreman sequence. For *Carphone* sequence the average PSNR gain of DBOMP-MSVR is observed to be 0.41 dB, 0.7 dB, and 0.95 dB as compared to BWT-MSVR, DBOMP-MLP, and BWT-MLP. The hybrid codec DBOMP-MSVR has a PSNR gain of 0.53 dB, 0.87 dB, and 1.43 dB over BWT-MSVR, DBOMP-MLP, and BWT-MLP for *Silent* sequence. Hence in general, the hybrid DBOMP-MSVR scheme shows improved RD performance as compared to other hybrid schemes in all considered video sequences.
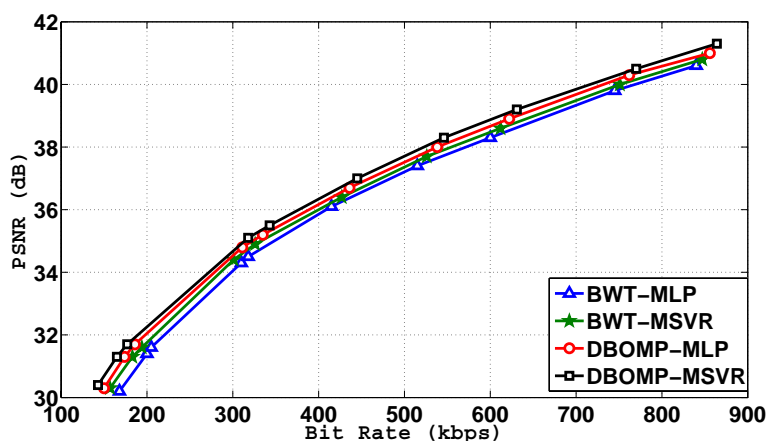
Figure 6.5: Overall rate distortion performance (*Foreman* at 30 fps)
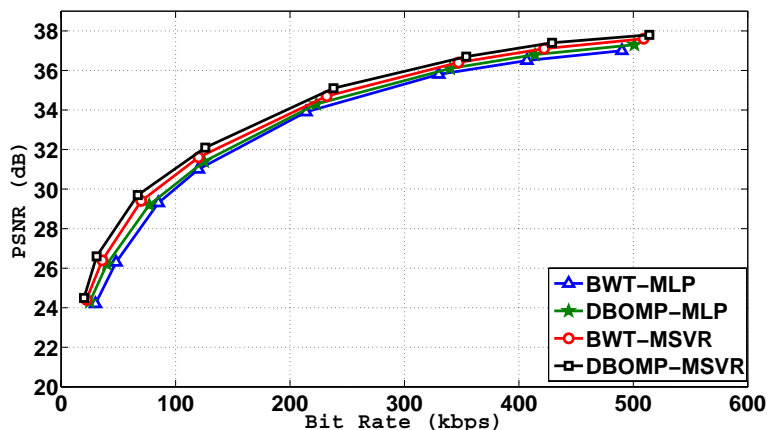
Figure 6.6: Overall rate distortion performance (*Carphone* at 30 fps)
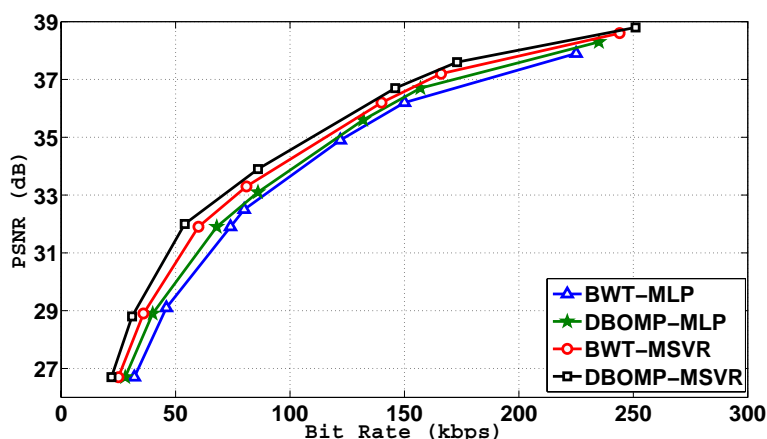
Figure 6.7: Overall rate distortion performance (*Silent* at 30 fps)

(b) **Analysis of number of requests per SI frame** :

This experiment evaluates the number of requests to be made in each frames during decoding. The frame number vs. the number of requests generated in different hybrid schemes for *Coastguard* and *Foreman* sequences are shown in Figures 6.8 and 6.9 respectively. For *Coastguard* sequence, the hybrid DBOMP-MSVR scheme has a maximum 763 requests in contrast to 780 in BWT-MSVR, 794 in DBOMP-MLP and 815 in BWT-MLP for frame number 30. Similarly, the maximum requests is 744 in DBOMP-MSVR and that in DBOMP-MLP is 716, 726 in BWT-MSVR, and 744 in BWT-MLP for frame number 110 for *Foreman* sequence. This comparison reveals the superiority of DBOMP-MSVR than the other hybrid schemes.

115
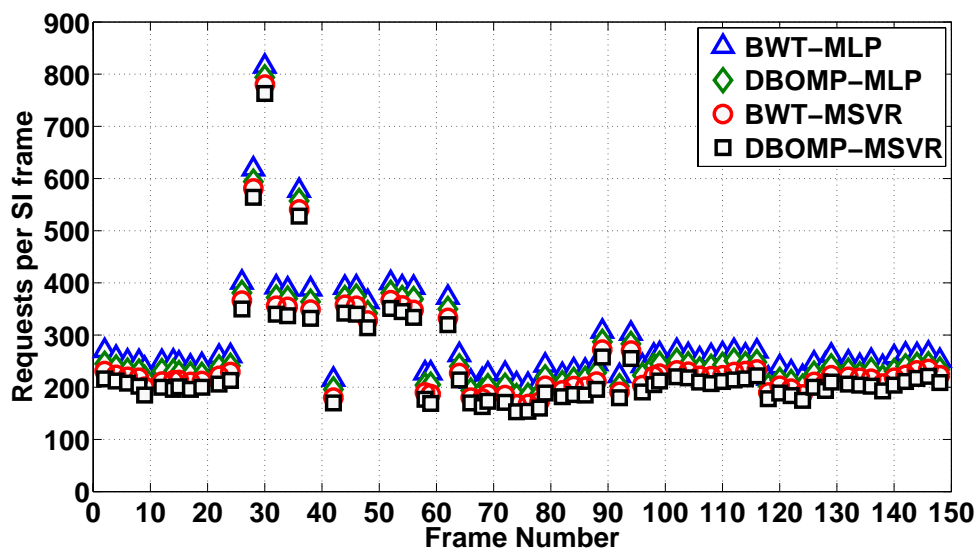
Figure 6.8: Number of requests per SI frame (*Coastguard* at 15 fps)
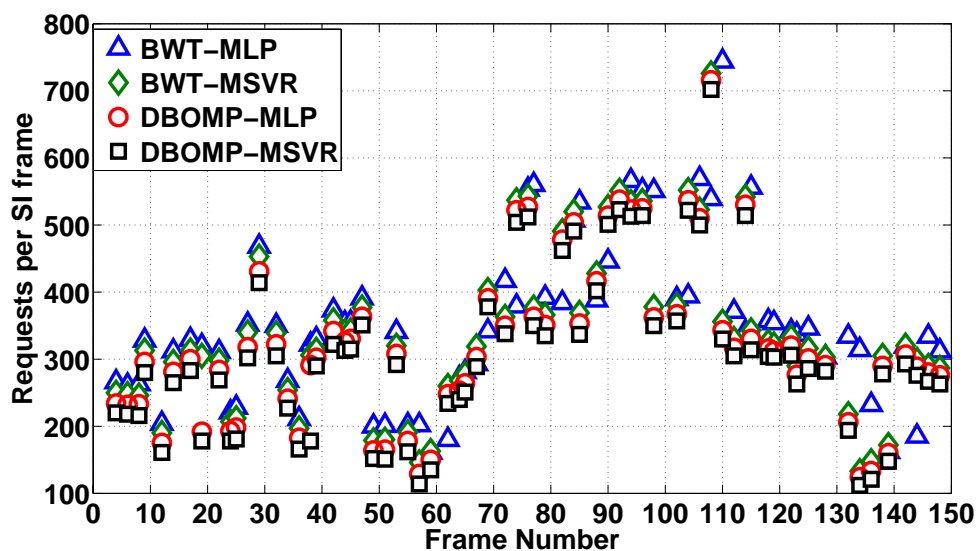
Figure 6.9: Number of requests per SI frame (*Foreman* at 15 fps)

(c) **Temporal evaluation of PSNR (dB)** :

To study the temporal frame quality, comparative analysis of PSNR (dB) is given in Figures 6.10 and 6.11 for *Foreman* and *Coastguard* sequences respectively. It may be observed that DBOMP-MSVR outperforms with respect to PSNR (dB) in both the cases. For computing the average PSNR for all the decoded frames, the hybrid DBOMP-MSVR scheme has gain of 0.76 dB, 1.07 dB and 1.47 dB over DBOMP-MLP, BWT-MSVR and

116

BWT-MLP schemes for *Foreman* sequence. For *Coastguard* sequence, the scheme DBOMP-MSVR has a gain of 0.64 dB, 0.93 dB and 1.26 dB over BWT-MSVR, DBOMP-MLP and BWT-MLP respectively.
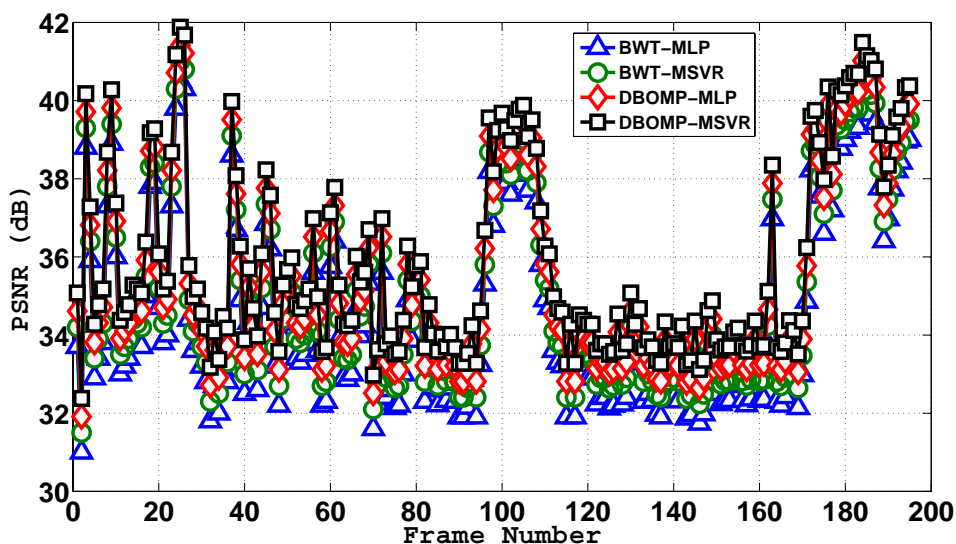
Figure 6.10: Temporal evaluation (*Foreman* at 30 fps)
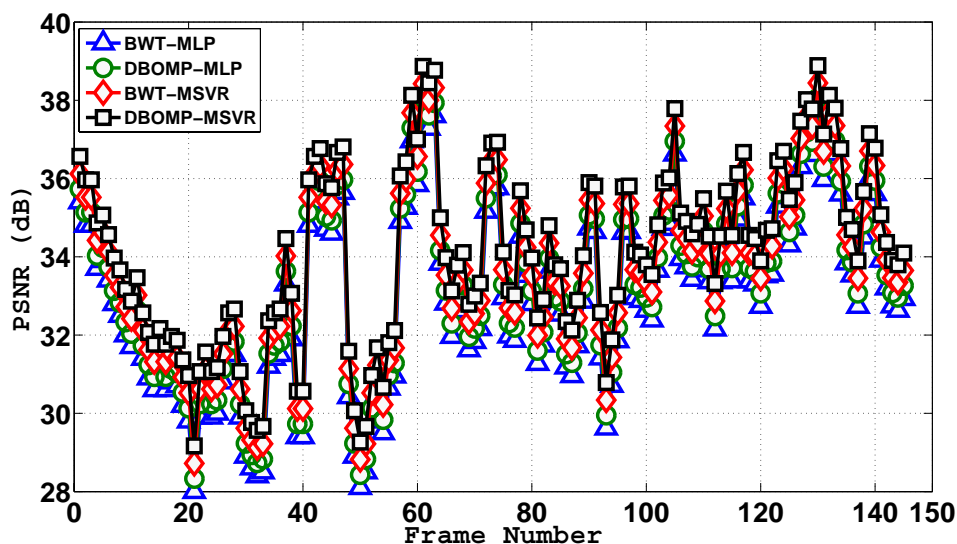
Figure 6.11: Temporal evaluation (*Coastguard* at 30 fps)

(d) **Analysis of decoding time requirement** :

To compare the decoding time requirement with different hybrid schemes, the decoding time (in seconds) are noted in each case and listed in Table 6.1.

117

A group of picture (GOP) size 2 and quantization matrices Q_4 and Q_8 are used uniformly for all the schemes. It is observed that the hybrid scheme DBOMP-MSVR adds no significant overhead as compared to other hybrid schemes. Rather DBOMP-MSVR scheme shows a reduction in decoding time due to generation of high quality key frames and WZ frames.

Table 6.1: Decoding time comparison of hybrid schemes

Sequences	Total decoding time(in seconds)			
	BWT-MLP	OMP-MLP	BWT-MSVR	OMP-MSVR
Foreman (Q_4)	1061	1032	1043	1021
Miss America (Q_4)	301	290	282	265
Coastguard (Q_4)	596	583	572	557
Foreman (Q_8)	3521	3487	3503	3476
Miss America (Q_8)	1117	1103	1084	1067
Coastguard (Q_8)	2914	2893	2881	2664

Summary :

The hybrid schemes generated from proposed intra-frame coding and side information generation schemes are evaluated through simulation. In general, it is observed that all the schemes have potential applications in DVC. Among all the schemes DBOMP-MSVR shows a superior performance as compared to other hybrid schemes.

Chapter 7

Conclusions and Future work

In this thesis, two schemes have proposed for intra-key-frame coding and two schemes for side information (SI) generation in a distributed video coding (DVC) framework. In DVC, intra-frame coding and side information are dependent on each other as SI uses decoded key frames. Hence superior quality key frames generated through intra-key frame coding in turn help in generating high quality SI frames. As a result, DVC needs less number of parity bits to reconstruct the WZ frames at the decoder. The suggested schemes are,

(a) Borrows Wheeler Transform (BWT) based H.264/AVC (Intra) intra-frame coding **(BWT-H.264/AVC(Intra)**)

(b) Dictionary based H.264/AVC (Intra) intra-frame coding using orthogonal matching pursuit (OMP) **(DBOMP-H.264/AVC (Intra))**

(c) Multilayer Perceptron (MLP) based side information generation **(MLP - SI)**

(d) Multivariable support vector regression (MSVR) based side information generation **(MSVR-SI)**

BWT-H.264/AVC (Intra) scheme is a modified version of H.264/AVC (Intra) scheme where a regularized bit stream is generated prior to compression. This scheme results in higher compression efficiency as well as high quality decoded key frames. **DBOMP-H.264/AVC (Intra)** scheme is based on an adaptive dictionary based H.264/AVC (Intra) intra-frame coding. The traditional

transform is replaced with a dictionary trained with K-SVD algorithm. The dictionary elements are coded using OMP. **MLP-SI scheme** utilizes a multilayer perceptron to estimate SI frames from the decoded key frames block-by-block. The network is trained offline using training patterns from different frames collected from standard video sequences. **MSVR-SI** scheme uses an optimized M-SVR to generate SI frames from decoded key frames block-by-block. Like MLP-SI, the training for M-SVR is made offline with known training patterns a priori. The suggested schemes are embedded into the Stanford based DVC architecture and studied individually to compare qualitative and quantitative performances with their competitive schemes. To exploit the usefulness of intra-frame coding schemes in SI generation, four hybrid schemes have been formulated in permutation of suggested schemes namely,

(a) **BWT-MLP** scheme that uses BWT-H.264/AVC (Intra) intra-frame coding scheme and MLP-SI side information generation scheme.

(b) **BWT-MSVR** scheme, where for intra-frame coding BWT-H.264/AVC (Intra) is used followed by **MSVR-SI** side information generation.

(c) **DBOMP-MLP** scheme is an outcome of putting DBOMP-H.264/AVC (Intra) intra-frame coding and **MLP-SI** side information generation schemes.

(d) **DBOMP-MSVR** scheme deals with DBOMP-H.264/AVC (Intra) intra-frame coding and **MSVR-SI** side information generation together.

The performance analysis of the hybrid schemes are also made with respect to overall rate distortion, number requests per frame, temporal evaluation, and decoding time requirement using standard video sequences to derive an overall conclusion. In general it is observed that the each scheme shows superior performance as compared to existing schemes and have a potential to apply in DVC architecture. Among the hybrid schemes, DBOMP-MSVR has an upper hand in all performance parameters. Further, the research directions are open to exploit other ANN schemes and optimization schemes for improved performance

of both intra-frame coding and side information generation. There exists a scope for improvement in other modules of DVC architecture.

Bibliography

Bibliography

[1] T. Sikora. MPEG digital video coding standards. *Signal Processing Magazine*, 14(5):82–100, September 1997.

[2] S. Park, Y. Lee, C. Kim, and S. Lee. CDV-DVC: Transform domain distributed video coding with multiple channel division. *Journal of Visual Communication and Image Representation*, 24(4):534–543, April 2013.

[3] Catarina Brites. Adavances on distributed video coding. Master's thesis, IST Portugal, December 2005.

[4] C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardi, and J. Ostermann. Distributed monoview and multiview video coding: basics, problems and recent advances. *IEEE Signal Processing Magazine, Special Issue on Signal Processing for Multiterminal Communication Systems*, pages 67–76, September 2007.

[5] B. Girod, A. Aaron, S. Rane, and D. Monedero. Distributed video coding. *Special Issue on Advances in Video Coding and Delivery*, 93(1):71–83, January 2005.

[6] R. Puri and K. Ramchandran. PRISM: a new robust video coding architecture based on distributed compression principles. In *Allerton Conference on Communication, Control and Computing*, pages 1–10, October 2002.

[7] Zhuo Xue. *Research and Developments of Distributed Video Coding*. PhD thesis, School of Engineering and Design, Brunel University, January 2009.

[8] J. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, July 1973.

[9] A. Wyner. Recent results in shannon theory. *IEEE Transactions on Information Theory*, 20(1):2–10, January 1974.

[10] A. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 22(1):1–10, January 1976.

[11] S. S. Pradhan and K. Ramchandran. Distributed source coding using syndromes (DISCUSS): design and construction. In *Proceedings of Data Compression Conference*, pages 158–167, March 1999.

[12] X. Wang and M. T. Orchard. Design of trellis codes for source coding with side information at the decoder. In *Proceedings of Data Compression Conference*, pages 361–370, May 2001.

[13] J. Bajcsy and P. Mitran. Coding for the Slepian-Wolf problem with turbo codes. In *IEEE Global Telecommunications Conference*, pages 1400–1404, April 2001.

[14] J. Garcia-Frias and Y. Zhao. Compression of correlated binary sources using turbo codes. *IEEE Communications Letters*, 5:417–419, October 2001.

[15] A.Liveris, Z.Xiong, and C.Georghiades. A distributed source coding using correlated images using turbo codes. *IEEE Communication Letter*, 6(9):379–381, September 2002.

[16] A. Aaron and B. Girod. Compression with side information using turbo codes. In *Proceedings of Data Compression Conference*, pages 252–261, October 2002.

[17] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero. Data compression of correlated nonbinary sources using punctured codes. In *Proceedings of Data Compression Conference*, pages 242–251, December 2002.

[18] Y. Zhao and J. Garcia-Frias. Joint estimation and compression of correlated nonbinary sources using punctured turbo codes. *IEEE Transactions on Communications*, 53(3):385–390, March 2005.

[19] P. Mitran and J. Bajcsy. Coding for the Wyner-Ziv problem with turbo like codes. *IEEE International Symposium Information Theory*, page 91, July 2002.

[20] J. Garcia-Frias and Y. Zhao. Compression of binary memoryless sources using punctured turbo codes. *IEEE Communications Letters*, 6(9):394–396, September 2002.

[21] G. C. Zhu and F. Alajaji. Turbo codes for non-uniform memoryless sources over noisy channels. *IEEE Communications Letters*, 6(2):64–66, February 2002.

[22] J. Garcia-Frias. Joint source channel decoding of correlated sources over noisy channels. In *Proceedings of Data Compression Conference*, pages 283–292, November 2001.

[23] A. D. Liveris, Z. Xiong, and C. N. Georghiades. Compression of binary sources with side information at the decoder using LDPC codes. *IEEE Communications Letters*, 6(10):440–442, October 2002.

[24] T. P. Coleman, A. H. Lee, M. Medard, and M. Effros. On some new approaches to practical Slepian-Wolf compression inspired by channel coding. In *Proceedings of Data Compression Conference*, pages 282–291, March 2004.

[25] C. F. Lan, A. D. Liveris, K. Narayanan, Z. Xiong, and C. Georghiades. Slepian-Wolf coding of multiple m-ary sources using LDPC codes. In *Proceedings of Data Compression Conference*, page 549, March 2004.

[26] N. Gehrig and P. L. Dragotti. Symmetric and asymmetric Slepian-Wolf codes with systematic and non-systematic linear coding. *IEEE Communications Letters*, 9(1):61–63, January 2005.

[27] X. Zhu, A. Aaron, and B. Girod. Distributed compression for large camera arrays. In *IEEE Workshop on Statistical Signal Processing*, pages 30–33, ST. Lewis, USA, September 2003.

[28] A. Aaron, S. Rane, E. Setton, and B. Girod. Tansform domain Wyner Ziv for video. In *Proceedings of SPIE Visual Communication and Image Processing*, pages 520–528, Sanjose,Carlifornia, January 2004.

[29] A. Aaron, S. Rane, and B. Girod. Hash based motion compensation at the receiver. In *Proceding of IEEE International Conference on Image Processing*, pages 3097–3100, Singapore, October 2004.

[30] R. Puri, A. Majumdar, and K. Ramchandran. PRISM: A Video Coding Paradigm with Motion Estimation at the Decoder. *IEEE Transactions on Image Processing*, 16:2436–2448, October 2007.

[31] A. Aaron, R. Zhang, and B. Girod. Wyner Ziv coding of motion video. In *Proceedings of Asilomar Conference on Signals and Systems*, pages 240–244, November 2002.

[32] R. Puri and K. Ramchandran. PRISM: a reversed multimedia coding paradigm. In *Proceedings of International Conference on Image Processing*, pages 617–620, September 2003.

[33] R. Puri and K. Ramchandran. PRISM: a new robust video coding architecture based on distributed compression principles. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 856–859, Hong Kong, China, 2003.

[34] A. Aaron, S. Rane, R. Zhang, and B. Girod. Wyner-Ziv coding for video: applications to compression and error resilience. In *Data Compression Conference*, pages 93–102, USA, March 2003.

[35] A. Aaron, E. Setton, and B. Girod. Towards practical Wyner-Ziv coding of video. In *Proceedings of International Conference in Image Processing*, pages 869–872, September 2003.

[36] C. Brites, J. Ascenso, and F. Pereira. Improving transform domain Wyner-Ziv video coding performance. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 525–528, May 2006.

[37] X. Guo, Y. Lu, F. Wu, and W. Gao. Distributed video coding using wavelet. In *Proceedings of International Symposium on Circuit and Systems (ISCAS)*, pages 5427–5430, May 2006.

[38] J. Ascenso, C. Brites, and F. Pereira. Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding. In *The 5th EURASIP Conf. Speech and Image Processing, Multimedia Communications and Services*, 2005.

[39] Y. Vatis, S. Klomp, and J. Ostermann. Inverse bit plane decoding order for turbo code based distributed video coding. In *International Conference on Image Processing*, pages 1–4, September 2007.

[40] J. M. Shapiro. Embedded image coding using zero trees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41(12):3445–3462, December 1993.

[41] A. Said and W. A. Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees (SPIHT). *IEEE Transactions on Circuits and Systems for Video Technology*, 6(3):243–250, June 1996.

[42] Y. Feng, D. Guiguang, D. Qionghai, and Y. Yaguang. Wavelet based scalable Wyner Ziv video coding. In *Proceedings of Picture Coding Symposium*, pages 13–19, September 2006.

[43] M. Grangetto, E. Magli, and G. Olmo. Context based distributed wavelet video coding. In *IEEE 7th Workshop on Multimedia Signal Processing*, pages 649–652, April 2006.

[44] B. Wu, X. Ji, D. Zhao, and W. Gao. Wavelet based distributed video coding with spatial scalability. In *IEEE International Symposium on Circuits and Systems*, pages 3458–3461, September 2008.

[45] Anhong Wang, Yao Zhao, and Hao Wang. Performance comparison of different channel codes in distributed video coding. In *IEEE International Conference on Image Processing*, pages 225–228, September 2006.

[46] S. Li, S. Fang, and Z. Li. Wyner-Ziv video coding for low bitrate using SPIHT algorithm. In *IEEE Workshop on Signal Processing Systems*, pages 341–345, October 2007.

[47] Y. Tonomura, T. Nakachi, and T. Fujii. Distributed video coding using JPEG 2000 coding scheme. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E90-A:581–589, May 2007.

[48] A. Wang, Y. Zhao, and J. S. Pan. An efficient hybrid distributed video coding. *IEICE Electronics Express*, 5:650–656, 2008.

[49] N. Cheung, C. Tang, A. Ortega, and S. Raghavendra. Effcient wavelet based predictive Slepian-Wolf coding for hyperspectral imagery. *Signal Processing*, 86(11):3180–3195, November 2006.

[50] A. Ponchet and Y. Iano. Distributed video coding using turbo codes and second generation wavelets. *IEEE Latin America transactions*, 6(7):565–571, December 2008.

[51] A. Wang, Y. Zhao, and H. Wang. LVQ based distributed video coding with LDPC in pixel domain. In *Lecture Notes in Computer Science*, pages 1248–1252, May 2006.

[52] J. Li, Y. Zhao, and A. Wang. Wavelet video coding combining pyramidal lattice vector quantization with Wyner-Ziv coding. In *Proceedings of International conference on signal Processing ICSP*, October 2007.

[53] B. Wu, X . Guo, D. Zhao, W. Gao, and F. Wu. An optimal non-uniform scalar quantizer for distributed video coding. In *IEEE International Conference on Multimedia and Expo*, pages 165–168, July 2006.

[54] F. Sheng, L. Xu-Jian, and Z. Li-Wei. A Lloyd-Max-based non-uniform quantization scheme for distributed video coding. In *The 8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel and Distributed Computing*, pages 848–853, July 2007.

[55] Q. Xu and Z. Xion. Layered Wyner-Ziv Video Coding. *IEEE Trans. Image Processing*, 15(12)(3791-3803), December 2006.

[56] F. Sheng, L.W. Zhang, and L. Huang. Adaptive nested scalar quantization in DVC. In *IEEE Workshop on Signal Processing Systems*, pages 351–356, October 2007.

[57] W.A.R.J. Weerakkody, W.A.C. Fernando, M.B. Badem, and A.M. Kondoz. Nonlinear Quantisation for Pixel Domain Distributed Video Coding. *IEEE Electronics Letter*, 45(5):1–2, February 2009.

[58] S. Subasingha and M.N. Murthi. On quantizer design for distributed source coding of Gaussian vector data with packet loss. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 14–19, March 2010.

[59] Y. Zhang and C.Zhu. Quantizer design for correlation noise in DVC. In *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pages 1–6, June 2011.

[60] R. Paresh and F. Lahouti. Multimode nested quantizer in presence of uncertain si and feedback. *IEEE Transactions on Communications*, 6(12):743–752, February 2013.

[61] D. Schonberg, S. S. Pradhan, and K Ramchandran. LDPC codes can approach the Slepian-Wolf bound for general binary sources. In *Presented at the AlIerton Conference Communication, Control, and Computing*, pages 1–5, 2002.

[62] G. Wu, L. Sun, and F. Huang. Consistent-quality distributed video coding framework. In *Lecture Notes in Computer Science*, pages 628–637, December 2007.

[63] R. P. Westeriaken, S. Borchert, R. K. Gunnewiek, and R. L. Lagendijk. Analyzing symbol and bit plane-based LDPC in distributed video coding. In *International Conference on Image Processing*, pages 17–20, September 2007.

[64] G. Xue, J. Xiaojun, and L. Zhang. Non-uniform LDPC for multi-bitplane in dvc. In *International Conference on Wireless Communication, Networking and Mobile Computing*, pages 1–4, September 2009.

[65] B. Li, Y. Wang, and Yu. Liu. Improved LDPC coding scheme with motion decision for DVC codec. In *Proceedings of IEEE LC-DNMT*, pages 540–544, October 2011.

[66] C. Brites, J. Ascenso, and F. Pereira. Augmented LDPC graph for DVC with multiple side information. In *Proceedings of International Workshop on Multimedia Signal Processing*, pages 1–6, October 2011.

[67] A. Aaron, D. Varodayan, and B. Girod. Wyner-Ziv residual coding of video. In *Proceeding of Picture Coding Symposium*, pages 1–5, March 2006.

[68] C. Brites, J. Ascenso, and F. Pereira. Modeling correlation noise statistics at decoder for pixel based Wyner-Ziv video coding. In *Proceedings of Picture Coding Symposium*, pages 16–21, September 2006.

[69] C. Brites, J. Ascenso, and F. Pereira. Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding. In *IEEE International Conference on Image Processing*, pages 273–276, October 2006.

[70] Catarina Brites and Fernando Pereira. Correlation noise modeling for efficient pixel and transform domain WynerZiv video coding. *IEEE Transaction on Circuits and Systems for Video Technology*, 18(9):177–1190, 2008.

[71] M. Dalai, R. Leonardi, and F. Pereira. Improving turbo codec integration in pixel-domain distributed video coding. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 537–540, May 2006.

[72] L. B. Qing, X. H. He, and Rui. LV. Modeling non-stationary correlation noise statistics for Wyner-Ziv video coding. In *International Conference on Wavelet Analysis and Pattern Recognition*, pages 316–320, November 2007.

[73] L. Qing, X. He, and Rui. Lv. Distributed video coding with dynamic virtual channel model estimation. In *Proceedings of First International Symposium on Data Privacy and E-Commerce*, pages 170–174, November 2007.

[74] M. Guo, Y. Lu, F. Wu, S. Li, and W. Gao. Distributed video coding with spatial correlation exploited only at the decoder. In *IEEE International Symposium on Circuits and Systems*, pages 41–44, May 2007.

[75] S. Fang, Z. Li, , and L. W. Zhang. Distributed video codec modeling correlation noise in wavelet coarsest subband. *Electronics Letters*, 43(23):1266–1267, November 2007.

[76] Y. Li, D. Zhao, and W. Gao. Modeling correlation statistics at decoder for multiview DVC. In *International Symposium on Circuit and Systems*, pages 2597–2600, May 2009.

[77] T. Tsai, L. Chang-Ming, and Y. Liu. Practical estimation of adaptive correlation noise model for DVC. In *International Symposium on Information Theory and its Applications*, pages 249–254, October 2010.

[78] Y. Vatis, S. Klomp, and J. Ostermann. Enhanced reconstruction of the quantized transform coefficients for Wyner-Ziv coding. In *Proceedings of International Conference on Multimedia and Expo (ICME), Beijing*, pages 172–175, 2004.

[79] D. Kubasov, J. Nayak, and C. Guilemot. Optimal reconstruction in Wyner-Ziv video coding with multiple side information. In *International Workshop on Multimedia Signal Processing*, pages 1–4, Greece, October 2007.

[80] B. Liu. Two pass reconstruction in DVC. In *International Conference on Multimedia Signal Processing*, pages 1–4, October 2009.

[81] B. Du and H. Shen. A novel reconstruction algorithm for DVC. In *International Conference on Future Computer and Communication*, pages 21–24, May 2010.

[82] M. Badem and W.A.C Fernado A. Kondoz. Unidirectional DVC using dynamic parity allocation and improved reconstruction. In *International Conference on Information and Automation for Sustainability*, pages 335–340, December 2010.

[83] L. Natario, C. Brites, J. Ascenso, and F. Pereira. Extrapolating side information for low-delay pixel domain distributed video coding. In *Lecture Notes in Computer Science*, pages 16–21, April 2006.

[84] X. Artigas and L. Torres. Iterative generation of motion-compensated side information for distributed video coding. In *IEEE International Conference on Image Processing*, pages 833–836, September 2005.

[85] A. Adikari, W. Fernando, and K. Arachchi. Multiple side information stream for distributed video coding. *IEEE Electronics letter*, 42(25):1447–1449, December 2006.

[86] A. Adikari, W. Fernando, H. Arachchi, and W. Weerakkody. Sequential motion estimation using luminance and chrominance information for distributed video coding of Wyner-Ziv frames. *IEEE Electronics letter*, 42(7):398–399, March 2006.

[87] W.Weerakkody, W. Fernando, J. Martinez, and F. Quiles P. Cuenca. An iterative refinement technique for side information generation in DVC. In *IEEE International Conference on Multimedia and Expo*, pages 164–167, July 2007.

[88] M. Badem, M. Mark, and W. Fernando. Side information refinement using motion estimation in dc domain for transform-based distributed video coding. *IEEE Electronics letter*, 44(16):965–966, July 2008.

[89] D. Varodayan, D. Chen, M. Flierl, and B. Girod. Wyner-Ziv coding of video with unsupervised motion vector learning. *Signal Processing Image Communication*, 23(8):369–378, June 2008.

[90] X. Artigas1, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret. The DISCOVER codec: architecture, techniques and evaluation. In

Proceedings of Picture Coding Symposium, pages 1–4, Lisbon, Portugal, November 2007.

[91] C. Brites, J. Ascenso, J. Pedro, and F. Pereira. Evaluating a feedback channel based transform domain Wyner-Ziv video codec. *Signal Processing Image communication*, 23(4):269–297, April 2008.

[92] J. Pedro., L. Ducla Soares, and F. Pereira. Studying error resilience performance for a feedback channel based transform domain Wyner-Ziv video codec. In *Proceedings of Picture Coding Symposium*, pages 1–5, Lisbon, Portugal, November 2007.

[93] R. Puri, A. Majumdar, P . Ishwar, and K Ramchandran. Distributed video coding in wireless sensor networks. *IEEE Signal Processing Magazine*, 23(4):94–106, July 2006.

[94] A. Adikari, W. Fernando, and W. Weerakkody. Independent key frame coding using correlated pixels in distributed video coding. *IEEE Electronics letter*, 43(7):387–388, March 2007.

[95] M. Burrows and D. Wheeler. A block sorting lossless data compression algorithm. Technical report, System research center, 130 Lyton Avenue, Palo Alto California, May 1994.

[96] P. Fenwick. The BWT block sorting text compression principle and improvements. *Computer Journal*, 39(9):781–790, September 1997.

[97] J. Abel. Improvements to the bwt compression algorithm after BWT stage. *ACM Transaction on Computer System*, 9(3):1–26, September 2003.

[98] A. Andersson and S. Nilsson. Implementing radix sort. *ACM Journal of Experimental Algorithmic*, 3:7–22, 1998.

[99] P. Fenwick. Block sorting text compression final report. Technical reports 130, University of Auckland, New Zealand, April 1996.

[100] H. Guo and C.S Burrus. Waveform and image compression using the Burrows Wheeler transform and the wavelet transform. In *IEEE International Conference on Image Processing*, pages 65–68, October 1997.

[101] S. Annadurai and R. Shanmugalakshmi. Color image compression using bit reduction and Burrows Wheeler transform. In *International Conference on*

Convergent technologies for Asia Pacific Region (TENCON), pages 895–898, October 2003.

[102] Y. Wiseman. Burrows-Wheeler based JPEG. *Data Science Journal*, 6:19–27, March 2007.

[103] E. Syahrul, J. Dubois, V. Vajnovszki, T. Saidani, and M. Atri. Lossless image compression using Burrows Wheeler transform methods and techniques. In *IEEE International Conference on Signal Image Technology and Internet Based Systems*, pages 338–343, May 2008.

[104] C. Wooten. *A Practical Guide to Video and Audio Compression*. ISBN: 0-240-80630-1. Elsevier, Focal Press, April 2005.

[105] T. Wiegend, G. Sullivan, and A. Luthara. Overview of the H.264/AVC video coding standard. *IEEE Transaction on Circuit and System for Video Technology*, 13(7):560–576, July 2003.

[106] G. Rath and C. Guillemot. On the simple derivation of complementary matching pursuit. *Signal Processing*, 90(2):702–706, 2010.

[107] G. Rath, C. Guillemot, and J.J Fuchs. Sparse approximations for joint source-channel coding. In *International Workshop on Multimedia Signal Processing*, pages 481–485, September 2008.

[108] R. Neff, A. Zakhor, and M. Vefferli. Very low bit rate video coding using matching pursuit. In *Proceedings of SPIE Conference in Visual Communications and Image processing*, pages 47–60, September 1994.

[109] R. Mallat and Z. Zhang. Matching pursuit with time frequency dictionaries. *IEEE Transaction on Signal Processing*, 41:3397–3415, December 1994.

[110] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decompositions. In *Proceedings of 27th Asilomar Conference on Signal, System and Compression*, pages 1–5, June 1993.

[111] M. Aharon, M. Elad, and A. Bruckstein. K-SVD : An algorithm for designing over complete dictionaries for sparse representaion. *IEEE Transaction on Signal Processing*, 54(11):4314–4322, November 2006.

[112] J. Ascenso and F. Pereira. Adaptive hash-based side information exploitation for efficient Wyner-Ziv video coding. In *International Conference on Image Processing*, pages 29–32, San Antonio, September 2007.

[113] M. Tagliasaccchi and S. Tubaro. Combining MCTF with distributed source coding. In *Proceedings of Visual Communication and Image Processing*, pages 797–800, Beijing, China, July 2005.

[114] D. Kubasov and C. Guillemot. Mesh-based motion-compensated interpolation for side information extraction in distributed video coding. In *IEEE International Conference on Image Processing*, pages 261–264, October 2006.

[115] T.N Dinh, G. Lee, J. Y. Chang, and H. Cho. Side information generation using extra information in distributed video coding. In *Proceedings of IEEE International Symposium on Signal Processing and Information Technology*, pages 138–143, December 2007.

[116] M. Tagliasacchi, S. Tubaro, and A. Sarti. On the modeling of motion in Wyner-Ziv video coding. In *International Conference on Image Processing*, pages 593–596, October 2006.

[117] V. Vapnik. *The Nature of Statistical Learning Theory*. ISBN-978-0-387-98780-4. Information Science and Statistics, Springer-Verlag, 2nd edition, 2000.

[118] F. Prez-Cruz, G. Camps E. Soria, J. Prez, A. R. Figueiras-Vidal, and A. Arts-Rodrguez. Multi-dimensional function approximation and regression estimation. In *Proceedings of International Conference on Artificial Neural Network (ICANN)*, Madrid, Spain, 2002.

[119] M. Sanchez, M. Cumplido, J. Garca, and P. Cruz. SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems. *IEEE Transactions on Signal Processing*, 52(8):2298–2307, August 2004.

[120] C. Wua, G. Tzeng, and R. Lin. A novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression. *Expert System with Application*, 36:22–31, August 2009.

[121] C. Huang and C. Wang. A GA-based feature selection and parameters optimization for support vector machines. *Expert System with Application*, 31:231–240, 2006.

[122] J. Kennedy and R. C. Eberhart. Particle swarm optimization. In *IEEE International Conference on Neural Networks*, pages 1942–1948, December 1995.

[123] J. Kennedy and R. C. Eberhart. A discrete binary version of the particle swarm algorithm. In *IEEE International Conference on Computational Cybernetics and Simulation*, pages 4104–4105, October 1997.

[124] H. Fan. A modification to particle swarm optimization algorithm. *Engineering Computations*, 19(8):970–989, October 2002.

Dissemination

Journals

1. **Suvendu Rup** and Banshidhar Majhi. A Mixed Framework for Transform Domain Wyner–Ziv Video Coding, *Elsevier Journal of Light and electron Optics (OPTIK)*, 124(21) : 4929-4938, 2013.

2. **Suvendu Rup** and Banshidhar Majhi. An Improved Side Information Generation for Distributed Video Coding. *Elsevier Journal of Electronics and Communication (AEU)*. Accepted for publication.

3. **Suvendu Rup**, and Banshidhar Majhi. Dictionary Based Intra-Frame Coding for Distributed Video Coding. *Submitted to ISA Transactions, Elsevier.*

4. **Suvendu Rup**, Banshidhar Majhi and Sudarshan Padhy. MSVR Based Side Information Generation with Adaptive Parameter Optimization in Distributed Video Coding. *Submitted to IEEE Transactions on Consumer Electronics.*

Conferences

1. **Suvendu Rup** and Banshidhar Majhi. Recent Advances in Distributed Video Coding . In *IEEE International Conference on Computer Science and Information Technology,* pages 130–135, Beijing, China, August 2009.

2. **Suvendu Rup** and Banshidhar Majhi. Burrows-Wheeler Transform Based Hybrid Video Coding. In *Second International Conference on Information and Communication Technology (IJCICT)*, pages 14-21 , Bhubaneswar, India, Jan 2010.

Suvendu Rup

Assistant Professor
International Institute of Information Technology, Bhubaneswar
Bhubaneswar – 753005, Odisha, India.

Ph: +91-674-3016062 (O), +91-9853125354 (M)

e-mail: suvendu@iiit-bh.ac.in

Qualification

- Ph.D. (Continuing)
  NIT Rourkela, Rourkela, Odisha

- M.Tech. (Computer Science & Engineering)
  Jadavpur University, Kolkata, West Bengal [First division]

- B.E. (CSE)
  Utkal University, Bhubaneswar, Odisha [First division]

Publications

- 02 Journal Articles

- 07 Conference Papers

Permanent Address

At – Swaroop Nagar, Ambika Sahi, Ward No – 05, Baripada – 753001, Odisha, India.

Date of Birth

July 05, 1978