Access to Electronic Thesis

Author:	Muhammad Majid
Thesis title:	Content scalability in multiple description image and video coding
Qualification:	PhD

CONTENT SCALABILITY IN MULTIPLE DESCRIPTION IMAGE AND VIDEO CODING

submitted by

Muhammad Majid

for the degree of

Doctor of Philosophy

of the

Department of Electronic and Electrical Engineering
The University of Sheffield

August, 2011

ABSTRACT

High compression ratio, scalability and reliability are the main issues for transmitting multimedia content over best effort networks. Scalable image and video coding meets the user requirements by truncating the scalable bitstream at different quality, resolution and frame rate. However, the performance of scalable coding deteriorates rapidly over packet networks if the base layer packets are lost during transmission. Multiple description coding (MDC) has emerged as an effective source coding technique for robust image and video transmission over lossy networks. In this research problem of incorporating scalability in MDC for robust image and video transmission over best effort network is addressed.

The first contribution of this thesis is to propose a strategy for generating more than two descriptions using multiple description scalar quantizer (MDSQ) with an objective to jointly decoded any number of descriptions in balanced and unbalanced manner. The distortion constraints and design conditions for multichannel unbalanced description coding (MUDC) using several MDSQs for improving quality as the number of description is increased are formulated. Secondly, the design of MDSQ is extended to incorporate the quality scalability in each description by using the concept of successive refinement in the side quantizers of multiple description scalar quantizer called MDSQ-SR. The design conditions of the MDSQ-SR are formulated with the objective to improve the quality of side and joint decoding for any combination of quality refinement layers. The joint decoding of different spatial resolution descriptions having different quality refinement layers is demonstrated for images by combining MUDC and MDSQ-SR schemes respectively. Finally, a fully scalable multiple description video coding (SMDVC) scheme is proposed by integrating MUDC and MDSQ-SR schemes in a motion compensated temporal filtering based video coding framework. The proposed SMDVC scheme is capable of generating and joint decoding any number of descriptions in balanced and unbalanced manner at any quality, resolution and frame rate.

According to the experimental results the unbalanced joint decoding results into 1.1 dB better peak to signal noise ratio (PSNR) than the balanced joint decoding at the same data rate. Furthermore, the joint decoding of MDSQ-SR based scheme gives an average of 1.35 dB and 0.3 dB better PSNR performance with respect to the state-of-the-art embedded-MDSQ for images and video respectively. The PSNR performance of the MDSQ-SR based video scheme is improved by 0.2-0.6 dB by controlling inter description and motion vector redundancies. In addition to superior rate-distortion performance than embedded-MDSQ, MDSQ-SR has reduced the computational complexity by 83%.

Dedicated to my grandmother

ACKNOWLEDGEMENTS

First and foremost I am thankful to Almighty ALLAH, without his blessings it is impossible to complete this thesis. My humblest and heartfelt Darood to Holy Prophet MUHAMMAD (Peace Be Upon Him), who educated his followers to seek the knowledge from cradle to grave. I would like to thank my supervisor Dr. Charith Abhayaratne who always gives me thought provoking ideas and useful suggestions throughout the PhD period. I feel myself lucky to have him as my supervisor who not only guides me on technical issues but also gives me the motivation and confidence to do research. Without his encouragement and patience, I could not finish my thesis so smoothly. I am grateful to Higher Education Commission (HEC) of Pakistan and University of Engineering and Technology (UET) Taxila, Pakistan for providing me the scholarship to pursue my higher education. I would like to thank Mr. James Screaton for the technical support. I am thankful to my colleagues in the lab for useful discussions. I would also like to thank my friends Anwar, Asif, Bilal, Hidayat, Nisar, Sohaib and Usman for sharing good times during my stay in Sheffield. I also want to thank my friends Naveed, Shahzad, Hammad, Imran and Asif. I am specially thankful to my brother Dr. Muhammad Sajid for his continuous support throughout my educational career. I would like to thank my sisters for their prayers and guidance. I am also thankful to my nephews and nieces for their love. Finally yet importantly, I wish to convey special thanks to my parents, without their prayers and support I am unable to reach this far with my studies.

Contents

List of Figures

List of Tables

List of Acronyms

2D	Two Dimensional
3D	Three DImensional
ARQ	Automatic Repeat Request
BMA	Block Matching Algorithm
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
EBCOT	Embedded Block Coding with Optimized Truncation
EMDSQ	Embedded Multiple Description Scalar Quantizer
EZW	Embedded Zero tree Wavelet
FEC	Forward Error Correction
GOP	Group of Pictures
IDCT	Inverse Discrete Cosine Transform
IF-MDVC	Independent Flow Multiple Description Video Coding
ITU	International Telecommunication Union
JPEG	Joint Photographic Expert Group
LOT	Lapped Orthogonal Transform
MC-EZBC	Motion Compensated Embedded Zero Block Coding
MCTF	Motion Compensated Temporal Filtering
MDC	Multiple Description Coding
MDSQ	Multiple Description Scalar Quantizer
MDSQ-SR	Multiple Description Scalar Quantizer with Successive Refinement
MDVC	Multiple Description Video Coding
MPEG	Moving Picture Expert Group
MSE	Mean Square Error
MUDC	Multichannel Unbalanced Description Coding
MV	Motion Vectors
NAL	Network Adaptation Layer
P2P	Peer to Peer
PCT	Pairwise Correlating Transform
PSNR	Peak Signal to Noise Ratio
ROI	Region of Interest
SDC	Single Description Coding
SSDVC	Single Scalable Description Video Coding
SMDC	Scalable Multiple Description Coding
SMDVC	Scalable Multiple Description Video Coding
SPIHT	Set Partitioning In Hierarchical Trees
SSIM	Structural Similarity Measure
SSQBM	Successive Side Quantizer Bin Merging
SVC	Scalable Video Coding
TDLT	Time Domain Lapped Transform

VCL	Video Coding Layer
VMR	Variance to Mean Ratio
VRC	Video Redundancy Coding
VQM	Video Quality Metric

List of Symbols

α	Magnitude shifting factor
δ	Central quantizer bin width
Δ	Implicit central quantizer bin width
γ	rate distortion region behaviour of multiple description coding
σ	Variance of Gaussian random variable
$\mathcal{T}$	Transformation matrix
ϕ	redundancy control parameter in pairwise correlating transform
$\mathcal{W}$	Gaussian random variable
$\mathcal{X}$	Gaussian random variable
$\mathcal{Y}$	Gaussian random variable
$\mathcal{Z}$	Gaussian random variable
A	Side information matrix
a	Central quantizer refinement factor
B	Number of group of pictures in video sequence
C	Central quantizer
c	Joint description decoding
D	Distortion
E	Number of temporal decomposition levels
f	Number of Diagonals filled in index assignment matrix
G	Number of spatial decomposition levels
g	Side quantizer bin spread
H	High frequency
I	Input image
$\hat{I}$	Decoded image
I_p	Index representation of side quantizer S for p refinement layers
h	Lead or lag factor of side quantizers
J	Number of multiple description scalar quantizers
J_p	Index representation of side quantizer T for p refinement layers
K	Execution time of coding block
L	Low frequency
l	Number of lost packet in each description
M	Total number of packets from each description
m	MDSQ index
N	Number of descriptions
O	Set of random variables
P	Number of refinement levels of side quantizers
p	Refinement level index
R	Rate requirement of description
r	Refinement factor of side quantizers

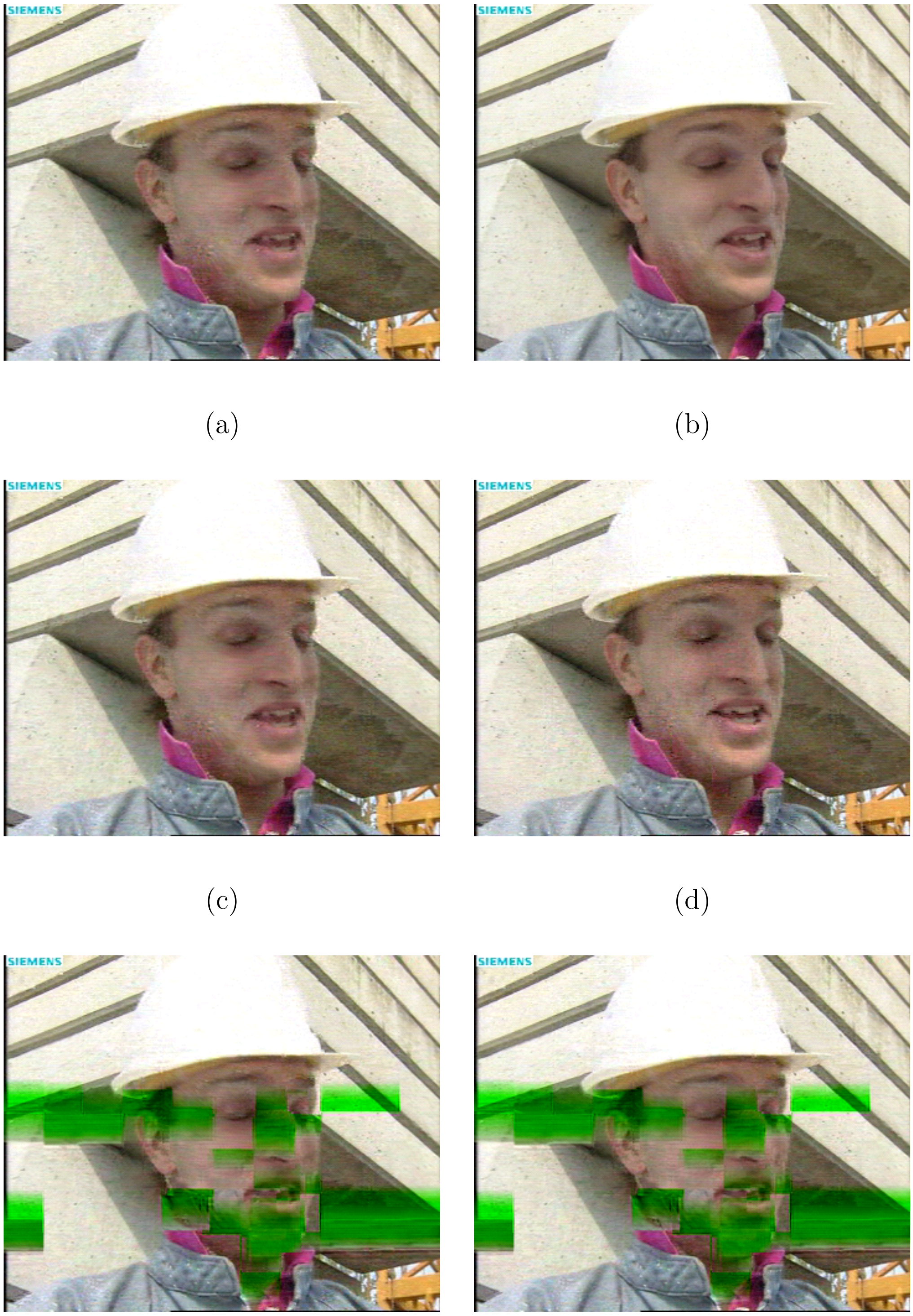
S	Side quantizer
s	Side description
T	Side quantizer
t	Side description
u	Minimum value of quantizer bin of side quantizer S
V	Input video sequence
$\hat{V}$	Decoded video sequence
v	Maximum value of quantizer bin of side quantizer S
W	spatio-temporal coefficients
w	Minimum value of quantizer bin of side quantizer T
X	Number of rows in video frame or image
x	Maximum value of quantizer bin of side quantizer T
Y	Number of columns in video frame or image
y	Minimum value of quantizer bin of central quantizer C
z	Maximum value of quantizer bin of central quantizer C

Chapter 1

Introduction

In spite of the advancement of multimedia content generation, communication and storage systems, there is a huge demand for efficient and robust image and video coding techniques for effectively utilizing the communication and storage resources. Conventionally image and video coders are designed to achieve high coding gain. But the emergence of the wavelet transform and other efficient hierarchical representations of the image and video data have resulted in incorporating scalability into image and video coders. Scalability is the property of the bitstream in which data is arranged according to the significance of information and can be truncated at any required rate, quality, resolution and frame rate. Therefore, scalable coding provides a good solution for image and video communication over heterogeneous packet switched networks where end users have different resources in terms of bandwidth and display.

The scalable coding is applicable for packet switched networks due to its adaptability to the heterogeneous environment. However, the scalable bitstream transmitted through network experiences packet losses, bit errors and transmission channel shutdowns. Packet losses and unavailability of communication path are the major problems encountered in multimedia content distribution over heterogeneous networks. The decoded quality of the image and video is deteriorated rapidly as the packets are lost from the base layer of a scalable bitstream or unable to decode in case of channel unavailability [1, 2]. In scalable coding, the higher enhancement layers are dependent on base and the lower enhancement layers.



(a) (b)

(c) (d)

(e) (f)

Figure 1.1: Decoded *Foreman* sequence frame from the scalable bitstream from base layer (left column) and base with enhancement layer (right column) having (a),(b) No packet losses, (c),(d) packet losses only in enhancement layer (e),(f) packet losses only in base layer.

Therefore, if the base layer is affected by transmission errors in error-prone channels, such errors are propagated due to interdependencies among layers and can lose the expected improvements in quality even though enhancement layers are received without any errors. Figure 1.1 shows the decoded frame from *Foreman* sequence for two quality layers. The left and right column of Figure 1.1 shows the decoded frame from the base layer and base with enhancement layer respectively. The decoded frame in Figure 1.1(a) and (b) are without any packet drops and clear quality improvement can be observed. The decoded frame in Figure 1.1(c) and (d) are with packet drop from the enhancement layer and it can be observed that the decoded quality remains similar as the base layer quality. The decoded frame in Figure 1.1(e) and (f) are with packet drops from the base layer and it can easily be observed from the decoded frame that by receiving the enhancement layers without any packet drops, the erasure effect in the base layer due to packet drop propagates to the enhancement layer. Therefore, some error resilient techniques are required to limit this type of error propagation.

The resilience to errors in error-prone channels is usually achieved by either using error control codes or error concealment techniques. Forward error correction (FEC) and automatic repeat request (ARQ) schemes have been commonly used for error correction in scalable multimedia transmission. The design of optimal forward error correction schemes is difficult especially for the best effort packet networks, where error occurs beyond the error correcting capabilities of the error correcting codes. On the other hand, automatic repeat request based methods give better performance than forward error correction schemes under packet lossy conditions, but the additional delays caused by the automatic repeat request schemes make them inappropriate for real time applications. Due to the limitations of the FEC and ARQ based schemes, a source coding scheme known as multiple description coding (MDC) has emerged as an effective scheme to combat channel impairments of heterogeneous packet networks [3–6]. MDC is feasible for real time applications because it, not only, combats packet losses without retransmission, but also, reduces the network congestion.

MDC is a well known source coding scheme that overcomes the drawbacks of FEC and ARQ methods for reliable transmission of images and video. In MDC, the source is encoded into two or more bitstreams, known as descriptions, and is transmitted independently through different channels. The descriptions can be decoded independently for a low quality version or jointly for a higher quality version of the same content. Independently decodable and mutually refineable descriptions are generated by introducing some amount of redundancy in each description, which is not useful if all the descriptions are received without any losses. If parts of individual description are affected from transmission losses, then the joint decoding compensates these losses and decodes the content as accurately as possible. MDC is useful for image and video transmission over best effort packet networks, where packets are lost due to various link bandwidths, buffer capacities and network congestion. MDC is also very applicable for distributed storage systems, where instead of saving similar content, different descriptions are stored on multiple servers.

In light of the above, it can be concluded that on one hand, image and video coding systems have to provide scalability in order to meet the requirements of the heterogeneous environment, and on the other hand, robust coding is also required to handle the challenges of the error-prone channels. Different methods have been proposed that combines the scalable and multiple description coding to provide both the scalability and reliability [7–12]. Most of the methods proposed for generating multiple scalable descriptions are based on single scalable coding schemes. In [9, 13, 14], scalable multiple descriptions are created by coding the temporal sub-sampled sequence from a single scalable video coder. In case of description loss in these schemes, the missing frames are replaced by interpolating the frames from the received description, resulting into ghosting effect on the side decoding. The other approach used for generating multiple scalable descriptions is to map the single scalable bitstream into multiple descriptions [8, 15–17]. In such schemes, the base layer information is duplicated in each description resulting in high redundancy and low joint quality gain. Also, most of the MDC methods focus on generating two descriptions. The methods for generating more than two descriptions available in literature are based on sub-sampling the source. The main problem of the sub-sampling based schemes is the severe effect on the reconstruction quality of the content if any of the description is not available at the decoder. In order to propose scalable multiple description image and video coding method the main aim and objectives set for this thesis are described in next section.

1.1 Aims and Objectives

In this thesis, research on scalable multiple description image and video coding that generates any number of descriptions and provides quality, resolution and frame rate scalability is presented. The main aim of this research is to design a scalable multiple description image and video coding method that generates any number of scalable descriptions which are capable of jointly decoding in balanced and unbalanced manner at any quality, resolution and frame rate provided that there is always an increase in the joint decoding quality. The main objectives set for this thesis are

1. To generate more than two descriptions using several multiple description scalar quantizers (MDSQs), with each description containing information from all coefficients as opposed to the subsampling based methods.

2. To make each description quality scalable by using the concept of successive refinement for the side quantizer of the MDSQ.

3. To provide the capability of joint decoding of different quality and resolution descriptions in multiple description image coding.

4. To integrate quality, resolution and frame rate scalability in a motion compensated temporal filtering based video coding framework.

5. To evaluate the proposed scalable multiple description video coding scheme under a practical scenario.

1.2 Thesis Outline

In Chapter 2 a brief overview of the building block of video coding and concepts of scalability and reliability is provided. Chapter 2 also reviews the state-of-the-art methods of multiple description image and video coding available in the literature.

In Chapter 3 a novel scheme based on MDSQ to generate and jointly decode any number of descriptions is presented. The joint decoding distortion constraints for different combination of descriptions from several MDSQs are formulated with the objective to improve the distortion as the number of jointly decoded descriptions is increased. For meeting these constraints, the design conditions for several MDSQs are proposed.

In Chapter 4 multiple description scalar quantizer with successive refinement (MDSQ-SR) is proposed to provide quality scalability in multiple description image coding. The proposed method starts with MDSQ based MDC for the base layer and then successively refine the side quanitzer. The objective of the MDSQ-SR design is to improve the distortion for every refinement layer of a side description when individually decoded and for any combination of levels of refinement of the two refined side descriptions for joint decoding. The MDSQ-SR design considers different index assignment matrices (resulting in non-overlapped and overlapped side quantizer bins) to incorporate different amounts of redundancy between the descriptions at the base layer.

Chapter 5 deals with generating different spatial resolution quality scalable descriptions to provide fully scalable multiple description image coding scheme. The proposed method generates different resolution descriptions by using multichannel unbalanced description coding concept for different wavelet decomposition levels and different quality enhancement layers is achieved in each description by successive refinement of the side quantizers.

Chapter 6 presents a new scalable multiple description video coding scheme. The temporal or frame rate scalability in the proposed scheme is achieved by decomposing the input video sequence by using motion compensated temporal filtering. The quality and resolution scalability is achieved by using MDSQ-SR and multichannel unbalanced description coding similarly as achieved in multiple description image coding. The chapter also deals with how each scalable description is extracted and how amount of redundancy in each description is controlled. Finally the proposed scalable multiple description video coding scheme is evaluated over peer to peer (P2P) network.

1.3 Author Publications

Some parts of the work presented in this research have been accepted or published as refereed journal and conference publications.

In Refereed Journal and Conference Proceeding

1. Muhammad Majid and Charith Abhayaratne, "Redundancy controllable scalable unbalanced multiple description bitstream generation for peer-to-peer video streaming", (accepted in) Signal Processing: Image Communication.

2. Muhammad Majid and Charith Abhayaratne, "Scalable multiple description video coding using successive refinement of side quantizers", in Proc. Picture Coding Symposium (PCS), 2010, pp. 602-605.

3. Muhammad Majid and Charith Abhayaratne, "Successive refinement of overlapped cell side quantizer for scalable multiple description coding", in Proc. Picture Coding Symposium (PCS), 2010, pp. 286-289.

4. Muhammad Majid and Charith Abhayaratne, "Distributed multiple description image coding", in Proc. IEEE International Workshop on Multimedia Signal Processing (MMSP), 2009, pp. 1-5.

5. Muhammad Majid and Charith Abhayaratne, "Fully scalable multiple description image coding", in Proc. IEEE International Workshop on Multimedia Signal Processing (MMSP), 2009, pp. 1-6.

6. Muhammad Majid and Charith Abhayaratne, "Multiple description scalar quantization with successive refinement", in Proc. EURASIP Signal Processing Conference (EUSIPCO), 2009, pp. 2268-2272.

7. Muhammad Majid and Charith Abhayaratne, "Multiple description image coding using several multiple description scalar quantizers", in Proc. SPIE Visual Communications and Image Processing, Vol. 7257, 2009, p. 725711 (9 pages).

Chapter 2

The Literature Review

Scalability and reliability in image and video coding are the two main components of this thesis. This chapter presents an overview of the concept of scalability and reliability in image and video coding systems. MDC is a source coding method that provides reliability for multimedia transmission over packet networks. This chapter also reviews the theoretical rate-distortion region of MDC and some practical MDC schemes for images and videos.

2.1 Building Blocks in Image and Video Coding

Significant amount of information is required to store any digital video sequence. Video coding or compression represents the video content by removing the redundant and irrelevant information present in a video sequence. The amount of data is reduced by exploiting interpixel, coding, and psychovisual redundancies present with in a frame or image. In addition to redundancies with in a frame, inter frame or temporal redundancy is reduced by coding some frames using motion compensated prediction with reference to previously coded frames. Video coding schemes can be lossy or lossless depending on the application. However, most of the video coding schemes are lossy to reduce the significant volume of data to represent the video. The distortion of the coded image or frame is usually

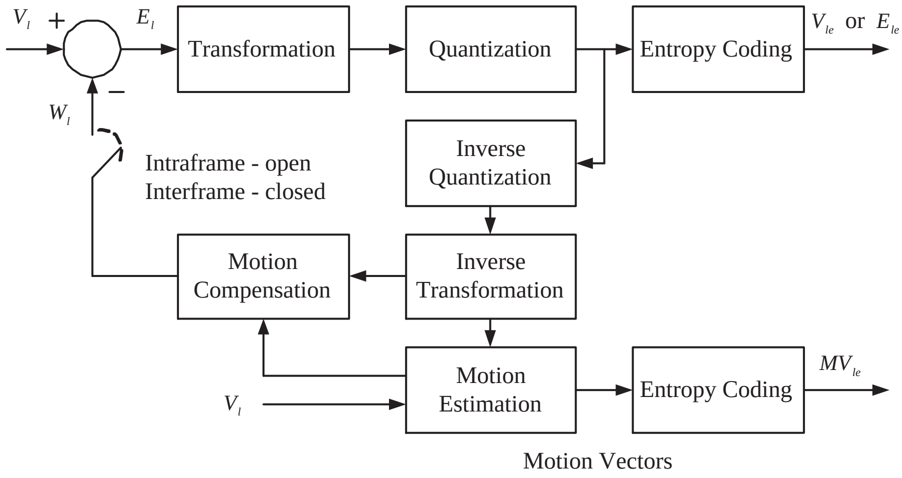


Figure 2.1: Block diagram of generalized video coding system.

measured by mean square error (MSE) or peak to signal noise ratio (PSNR) as,

$$MSE = \frac{1}{X \times Y} \sum_{i=1}^{X} \sum_{j=1}^{Y} [V(i,j) - \hat{V}(i,j)]^2, \tag{2.1}$$

$$PSNR = 20log_{10} \left(\frac{255}{\sqrt{MSE}} \right) dB, \tag{2.2}$$

where V,$\hat{V}$ are the original and decoded video frame and $X \times Y$ is the dimension of frame.

Figure 2.1 shows the block diagram of a generalized video coding system that operates in intraframe and interframe video coding modes to remove redundancies with in the frame and among different frames respectively. The intraframe video coding mode exploits the redundancies present within the frame as in image coding. Therefore, no feedback loop is required in this mode as shown in Figure 2.1. Each frame V_l is decorrelated by a transformation block and then quantized and entropy coded to reduce the bit budget. Inverse operations are performed on encoded intraframe V_{le} to get the reconstructed frame $\hat{V}_l$. On the other hand, the interframe video coding exploits the dependencies among different frames to increase the coding efficiency.

In interframe video coding, the first frame of each group of picture (GOP) is coded as intraframe and then the motion prediction based coding is used on the following frames of each GOP. The reason of coding the first frame as intraframe is to prevent the coder from error propagation. In interframe coding mode, the prediction of current frame W_l is subtracted from current frame to form the prediction error E_l as shown in Figure 2.1. The current prediction error is encoded similarly as in intraframe mode by performing transform, quantization and entropy coding with entropy coded motion vectors MV_l. The decoder uses the encoded prediction error E_{le}, encoded motion vectors MV_{le} and previous reconstructed frame $\hat{V}_{l-1}$ to reconstruct the current frame $\hat{V}_l$. The current frame prediction is generated at the decoder by using previous reconstructed frame and motion vectors. The current frame reconstruction is then obtained by adding reconstructed prediction error and current frame prediction from motion vectors. Details of each block in intraframe and interframe video coding modes are as follows.

2.1.1 Transformation

The main purpose of the transformation block is to remove the interpixel redundancy present within the frame. The transformation block decorrelates the input frame by distributing most of the input frame energy into fewer numbers of coefficients. The transformation operation in image and video coding is considered as a lossless process because the transforms used in these coding are invertible. Therefore, the original frame can be reconstructed by performing inverse transformation operation. The transform used in coding algorithms can be applied to the entire frame or blocks in the frame. The most common block-based transform used in image and video coding is the discrete cosine transform (DCT). The DCT is usually performed on 8×8 blocks, therefore well suited for fast real time implementations, but produces blocking artefacts in coded frames.

The discrete wavelet transform (DWT) is another transform used in image and video coding and it shows superior performance over DCT-based methods [18]. The DWT can be implemented by using the filter bank or the lifting approach. The 2D-DWT decompose the frame into four (LL, LH, HL, HH) subbands, where L and H stand for low and high frequency respectively. The LL subband represents the half resolution of the original frame with high spatial correlation and

LH, HL and HH subbands represent the vertical, horizontal and diagonal edge details present in the frame. The DWT can be applied to subbands to further decompose the input frame. The wavelet decomposition structures can be dyadic, in which only LL subband is further decomposed, or packet transform where all subbands are further decomposed. Spatial scalability can easily be achieved in DWT-based coding due to its multi-resolution decomposition structure.

2.1.2 Quantization

The main purpose of the quantization block is to remove the psychovisual redundancy. The quantization is performed on the transformed coefficients and is a lossy process, thus results in high coding gain. The objective of quantization is to generate a finite number of symbols which are the approximation of the transformed coefficients. The quantizer used in image and video coding can be scalar or vector. In vector quantization, the transformed coefficients are divided into blocks and then the quantization symbols are assigned to each block [19]. On the other hand, in scalar quantization each transformed coefficient is mapped to one quantization symbol according to the coefficient value and its correspondent quantizer bin. The scalar quantizer used in coding system can be uniform or non-uniform and is selected according to the rate-distortion requirement of the coding system.

2.1.3 Entropy Coding

The main purpose of the entropy coding block is to exploit the coding redundancy present in the quantized coefficients. Before transmission, the quantized coefficients are represented losslessly in terms of binary stream. The binary codewords assigned to quantized coefficients can be of fixed or variable length. Compression ratio can be increased by using variable length codewords. The two famous entropy coding methods are Huffman [20] and arithmetic coding [21]. In Huffman coding, integer length codeword is assigned to each message symbol. Therefore, the bit rate cannot be less than one bit per message symbol, unless the message symbols are coded jointly. On the other hand in arithmetic coding, each message symbol does not need to be mapped into integer number of bits. Therefore, fractional bit rate can be achieved in arithmetic coding.

2.1.4 Motion Compensation and Estimation

The main purpose of the motion compensation and estimation block is to remove interframe redundancy present in the temporal direction of video sequence. Motion compensation is a technique used to remove the temporal redundancy that also enhances the coding efficiency of video encoders [22]. The motion compensation-based video coders work in two stages. In the first stage, motion is estimated between two frames *i.e.,* the current and previously reconstructed frame. Block matching algorithms (BMA) [23] are usually used to estimate the motion between two frames. The motion estimator generates motion vectors by dividing the current frame into blocks and then each block is estimated from the search window of the previously reconstructed frame. The complexity of the BMA depends on the block size (that can be of fixed or variable size) and search window size. One of the criteria used to choose the best block match between frames is MSE.

In the second stage, the current frame prediction is created by using the motion vectors generated from motion estimation block and previously reconstructed frame. The blocks from the previously reconstructed frame are placed according to the motion vectors to create the current frame prediction. The motion compensation also decides which blocks are encoded as prediction error and which are encoded similarly as in intraframe coding. On the decoder side, the frames are reconstructed from the entropy encoded motion vectors and transformed, quantized and entropy coded prediction error.

2.1.5 Image and Video Coding Standards

In the literature, different image and video coding standards are proposed for different applications based on different transform, quantization, motion compensation and entropy coding. Any coding standard explains the bitstream structure which describes the building blocks configuration in image and video coding system and other information to represent images and videos.

The joint photographic expert group (JPEG) proposed the compression standard to represent still images in which different modes can be used to fulfil the requirements for particular application. The JPEG baseline is a simple and efficient system that uses the DCT-based lossy compression algorithm with scalar quantization and Huffman or arithmetic coding to represent an image in a non scalable fashion, also called sequential mode. The JPEG extended system enhances the baseline system by incorporating scalability in progressive and hierarchical modes. In the progressive mode, each DCT coefficient block is roughly quantized and then refined sequentially to achieve quality scalability. On the other hand in the first pass of hierarchical mode, the subsampled image is coded by JPEG baseline. The difference between the higher resolution image and prediction obtained from the upsampled and interpolated compressed image from the previous passes are coded in subsequent passes. JPEG standard also provides a lossless coding mode that uses a predictive coding approach with Huffman or arithmetic coding and is independent from DCT. The details of the JPEG standard can be found in [24].

Despite the success of JPEG image coding standard, it has certain shortcomings for different applications like medical imaging, digital libraries and archives of images, and communication of images over Internet and mobile. JPEG2000 image coding standard not only optimizes the coding efficiency but also provides the scalability and interoperatability for Internet and mobile communication. In JPEG2000 block-based DCT is replaced by the DWT, which not only enhances the compression efficiency but also represents the image in multi-resolution form. In addition to high coding gain, JPEG2000 also provides different features like lossy and lossless coding modes, quality and resolution scalability and region of interest (ROI)-based coding. The details of JPEG2000 and comparison with other image coding standards are reviewed in [25, 26].

Different video coding standards are available, considering different application requirements. The moving picture expert group (MPEG) has developed MPEG-1, MPEG-2, MPEG-4, MPEG-7 and MPEG-21. The MPEG-1 video standard supports CIF format at 25 fps or 30 fps, non-interlaced video to encode up to 1.5

Mbps. MPEG-1 is designed for CD ROMs and multimedia application for the desktop computers. MPEG-2 video coding standard is an extension of MPEG-1 that supports interlaced video coding. It is used for standard and high definition television broadcast over satellite, terrestrial emission and cable networks and for high quality video storage on DVDs. MPEG-2 is also used for telecommunication purposes and is recommended by international telecommunication union (ITU) as ITU-T H.262. The H.263 video coding standard is developed for low data rate applications, especially for mobile wireless networks. The main aim of any standard is to maximize the coding gain, to reduce the complexity of coder, and to make them applicable for diverse networks. The H.264/AVC covers video coding layer (VCL) and network adaptation layer (NAL) to provide flexibility and customization to applications. VCL is responsible to represent the video content effectively, and NAL is responsible for header information to pack the VCL data for network transport. The details of the H.264/AVC and its scalable extension is briefly reviewed in [27–29].

2.2 Scalable Image and Video Coding

The main aim of conventional image and video coders is to achieve high compression ratio or coding gain. But high compression ratio is not always the only requirement of the end user, especially when the end users have different resources in terms of bandwidth, display device and computational complexity. Scalable coding emerges as a good solution for multimedia content distribution over heterogeneous networks.

Hierarchical subband decomposition and embedded coding are the two main components of any scalable coding framework. Scalability is the property of a bitstream in which the bitstream is arranged according to the significance of information and can be truncated. Therefore the scalable image and video codecs allow the end users to truncate the scalable bitstream at any frame rate, resolution and quality to meet the data rate requirement and user preferences. Any scalable coder generates an embedded bitstream and has at least two layers *i.e.,* base layer and enhancement layers. The base layer contains the most important information by which a minimum quality or resolution is obtained [26]. The base layer is followed by other layers, called the enhancement layers, having additional information to enhance the quality, resolution or frame rate of the decoded image or video. Following are the different types of scalabilities that are useful in image and video coding.

1. **Quality or SNR Scalability:** In quality or SNR scalability, at least two layers (base and enhancement) of an image/video are required to decode the image/video at two or more different quality levels. The base layer encodes the information that is required to decode the image/video at a basic quality. The enhancement layer increases the quality of the decoded image/video when added to the base layer. The Encoder can encode as many enhancement layers as possible which gives decoder an option to decode the image/video at different quality levels.

2. **Spatial or Resolution Scalability:** In spatial or resolution scalability, the base layer generated by the encoder is responsible to provide a basic lower spatial resolution. The enhancement layer provides the information, which is interpolated with the base layer to decode the image at some higher spatial resolution.

3. **Temporal or Frame Rate Scalability:** In temporal scalability different frame rates can be selected for video encoding/decoding. Fewer frames from the video sequence are used for the motion prediction and estimation for the base layer. Higher frame rate are used in the enhancement layer for the good perception of motion in video.

The quality and resolution scalabilities can be achieved both in images and video while the temporal scalability is possible only in videos. Figure 2.2 shows the scalable video coding framework. A scalable video coding framework is divided into three main blocks [1, 30].

1. Scalable Video Encoder.

2. Scalable Video Extractor.

3. Scalable Video Decoder.

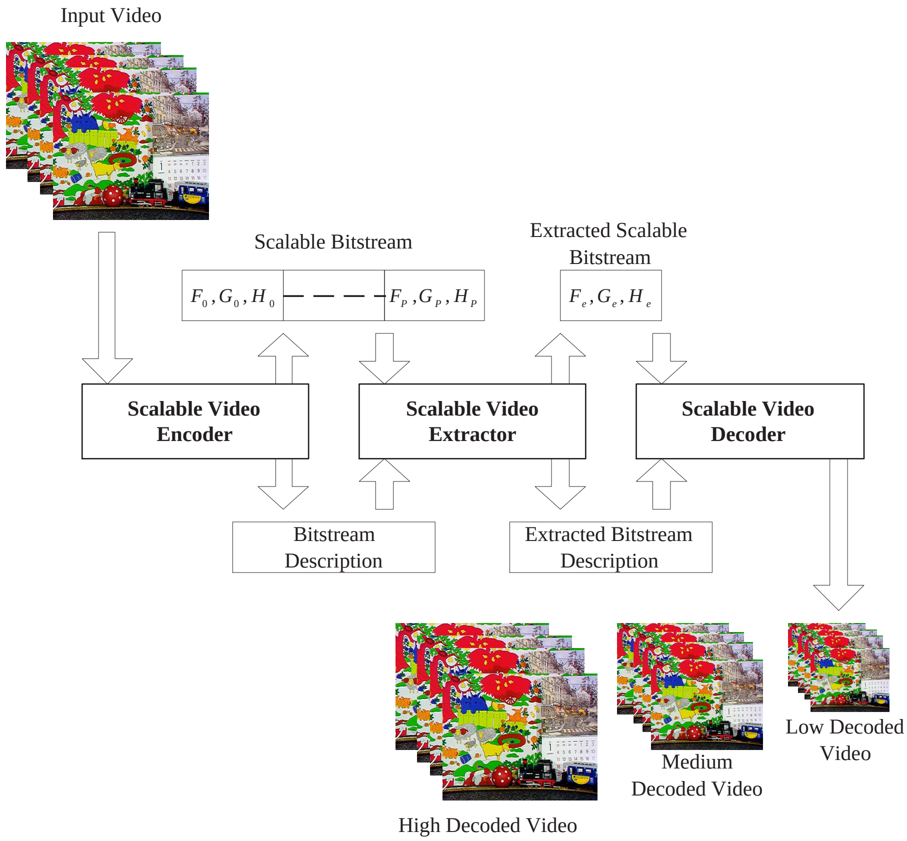


Figure 2.2: Scalable video coding framework.

The encoder block only once generate a scalable video bitstream and a bitstream description for input video for the highest achievable quality, resolution and frame rate. The bitstream description can be used separately or interleaved with the scalable video bitstream. The scalable video bitstream is generated in such a manner, that it is capable of achieving all the three types of scalabilities discussed above. The extractor block is responsible to truncate the scalable video bitstream into a new adapted scalable video bitstream and its description. The decoder block uses the adapted scalable video bitstream and its description to decode the input video at particular quality, resolution and frame rate depending on the adapted scalable video bitstream.

Let F_0, G_0 and H_0 be the bitstream requirement for a basic quality, resolution and frame rate respectively and F_P, G_P and H_P be the bitstream requirement for a highest quality, resolution and frame rate as shown in Figure 2.2. All this information is presented in a single bitstream generated by the scalable video encoder. Extractor can extract the scalable video bitstream at any quality (F_e), resolution (G_e), and frame rate (H_e). Decoder decodes the input video at a different quality, resolution, and frame rate according to extracted scalable bitstream and its description.

Different scalable image coding algorithms are available in the literature. Shapiro in [31] introduced the concept of embedded zero tree wavelet (EZW)-based image coding that generates a bitstream according to the significance of the wavelet coefficients. An alternative scheme for implementing the same concept as introduced in EZW is discussed in [32] named SPIHT (Set Partitioning in Hierarchical Trees). Only quality scalability is achieved in EZW and SPIHT. Both the quality and resolution scalability is achieved in embedded block coding with optimized truncation (EBCOT) [33], which is also adopted in JPEG2000 [26]. Motion compensated temporal filtering (MCTF) [34] and DWT is extensively used in video coding to generate scalable video bitstreams [27, 35–40]. MCTF is a lifting based wavelet approach used to decompose a video in temporal direction. Motion compensation and prediction is performed in [41, 42] and not performed in [43] when applying the wavelet transform in temporal direction.

3D wavelet decomposition or spatio-temporal decomposition is a two step process: 2D spatial transform and MCTF. In video coding two different frameworks for the spatio-temporal decomposition are used. In one framework, MCTF is performed on 2D spatial transform coefficients and is known as (2D+t) framework [44]. In another framework, the 2D spatial transform is performed after the MCTF and is known as (t+2D) framework [2]. All three kinds of scalabilities *i.e.,* (temporal, spatial and quality) can be achieved by using the spatio-temporal decomposition architecture. Motion vectors generated by MCTF can be encoded in non scalable fashion in [2] and also in scalable fashion [45, 46]. In [1], different wavelet-based scalable video coding approaches are discussed in detail.

The major problem of scalable bitstream is its rapid performance deterioration when transmitted over error-prone channels. In scalable coding, the higher enhancement layers are dependent on base and the lower enhancement layers. Therefore, if the base layer is affected by transmission errors in error-prone channels, such errors are propagated due to interdependencies among layers and can lose the expected improvements in quality even though enhancement layers are received without any errors. An example of such a situation is the best effort packet networks like Internet. The scalable video bitstream is packetized according to the significance of information to transmit the video through Internet. If the packets are corrupted or lost at different nodes due to various bandwidth links, buffer capacities and network congestion, it is possible that the video is not decoded properly at the decoder. In such cases some error resilient image and video coding methods are required to cater for the error propagation problem of scalable bitstreams.

Different methods have been proposed for reliable transmission of images and video over Internet and mobile wireless networks. Forward error correction and automatic repeat request are the two common error correction techniques adopted in image and video transmission over error-prone channels [47,48]. The data rate is increased by introducing any error correction technique. FEC schemes are capable to detect and correct certain amount of bit errors depending on the error detection and correction capabilities of the adopted schemes. The forward error correction scheme fails when the bit errors are beyond the error correction capabilities. Usually these schemes fail under bursty error conditions. It is shown in [49, 50] that the ARQ is more effective to combat bursty errors than the FEC scheme. An additional delay caused by the ARQ scheme for requesting the corrupted packets is only the disadvantage and therefore it is not appropriate for real time applications. Instead of using FEC or ARQ a source coding method known as MDC is also used as an effective scheme to combat channel errors.

2.3 Multiple Description Coding

MDC has emerged as a source coding technique that overcomes the drawbacks of FEC and ARQ schemes for reliable transmission over unreliable channels. In contrast to scalable coding, no error correction or concealment techniques are required in MDC for robust transmission. MDC provides graceful degradation as compared to the conventional scalable coders because in MDC, the source information is split into two or more descriptions and is transmitted through different independent channels. In MDC, the source is encoded into different bit streams with similar rate-distortion performance known as balanced descriptions, which can be decoded independently for a low quality version or jointly for a higher

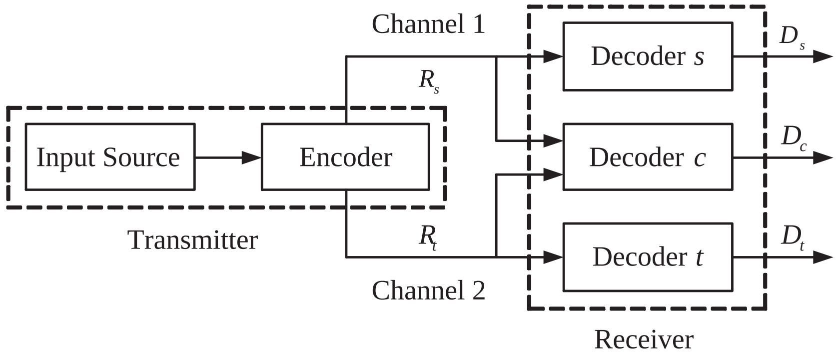


Figure 2.3: Multiple description coding model for two descriptions.

quality version of the same content [3]. If parts of the individual descriptions are affected from transmission losses, then the joint decoding compensates for these errors and decodes the content as accurately as possible. MDC is useful for transmission along packet networks, where loss of packets occur due to various link bandwidths, buffer capacities and network congestion and for transmission along wireless channels, where bit wise errors occur due to fading. More importantly, MDC is very applicable in distributed and scalable content storage and transmission systems [51–54].

The block diagram of a MDC model for two descriptions is shown in Figure 2.3. An input source is transmitted to three decoders over two different channels. On the receiver side, one decoder receives the information from both the channels and is called a central decoder. The other two decoders receive the information from independent single channel and are called the side decoders.

Let s and t be the two descriptions having data rate requirements R_s and R_t and are transmitted through channel 1 and channel 2 respectively. Decoder s and t in Figure 2.3 are the side decoders, while decoder c is the central decoder. Different distortion levels on the decoder side can be achieved depending on the rates and the number of descriptions received. On receiving a single description at the decoder, one of the side decoders decodes the source information. The level of distortion is either D_s or D_t depending on which of the side decoder receives the description. When both the descriptions are received, the distortion level is D_c, which should be less than both the side distortions at rate R_c. For the balanced descriptions case the above conditions can be written as

$$\begin{aligned} D_s &\approx D_t, \\ R_s &\approx R_t, \\ D_c &\leq D_s, D_t, \\ R_c &= R_s + R_t. \end{aligned} \tag{2.3}$$

The above MDC model can be generalized to N descriptions, in that case there are N channels and $2^N - 1$ decoders.

It is not always necessary to create balanced descriptions in terms of rate and distortion. Three very simple and different scenarios to create multiple descriptions are discussed in [3]. In first scenario, two balanced but different descriptions are generated at rate R_s and R_t and transmitted through two channels. In this case, a minimum quality level decoding is achieved by the side decoders, while a maximum level of quality is achieved at the central decoder. In second scenario, the same description is transmitted through both the channels. In this case the decoding quality achieved by the side and central decoders is same. There is no advantage at the central decoder under lossless condition because no additional information is available at the central decoder. In third scenario, description s is created at R_s and description t is created at rate R_t in such a way that it has some enhancement information regarding to description s. In this case, a good quality decoding is achieved by the side decoder s and even a better quality is achieved at the central decoder. But the decoding quality achieved by the side decoder t is not acceptable because the description t just has the enhancement information of the description s. Therefore, the fundamental tradeoff of any MDC scheme is to make the description individually good but not too similar [3].

Different methods of MDC have been proposed by integrating the creation of multiple descriptions into image and video encoding modules, such as, the decorrelating transform and quantization. Another MDC approach is to create different spatio-temporal versions of the source content by subsampling followed by individual encoding of each of the description using existing source coders. The next sections of this chapter reviews the theoretical rate-distortion region of MDC and some practical MDC schemes for images and videos.

2.4 Rate Distortion Region for Multiple Description

In this section, the MDC problem from the information theory perspective is reviewed. Different information theoretic aspects of MDC problem regarding rate and distortion are studied that helps in designing practical MDC schemes [55–59]. It is always difficult to find the tight rate and distortion bound except for some simple situations. However, the rate and distortion bound helps in understanding the quality variation with respect to the source coding length. For single description coding, the rate-distortion pair (R, D) is achievable if a source code exists at rate R to represent the source having distortion D. Similarly, the rate distortion region is a closure of the set of achievable rate-distortion pairs. Most of the rate distortion region studies on MDC is for a classical two description case, where source is encoded into two descriptions at rate R_s and R_t. The decoder can have the distortion D_s or D_t and D_c depending on single and joint description decoding respectively. The multiple description rate distortion region is a closure set of achievable quintuples $(R_s, R_t, D_c, D_s, D_t)$. The multiple description rate distortion region is only known for memoryless Gaussian source and mean squared error and is discussed in next section.

2.4.1 Rate Distortion Region of Memoryless Gaussian Source and Mean Squared Error Distortion

The achievable multiple description rate distortion region is completely known only for a memoryless Gaussian source and the result is presented by Ozarow in [56]. In [56], a source of sequence of independent and identically distributed random variables having Gaussian distribution with unit variance is considered. The achievable set of rates and mean square error distortion is the union of points that satisfies the following equations.

$$D_s \geq 2^{-2R_s}, \tag{2.4}$$

$$D_t \geq 2^{-2R_t}, \tag{2.5}$$

$$D_c \geq 2^{-2(R_s+R_t)}\gamma(D_s, D_t, R_s, R_t), \tag{2.6}$$

where,

$$\gamma(D_s, D_t, R_s, R_t) = \frac{1}{1 - \left(\sqrt{(1-D_s)(1-D_t)} - \sqrt{D_s D_t - 2^{-2(R_s+R_t)}}\right)^2}. \tag{2.7}$$

The behavior of γ and the properties of achievable region of MDC is made clear by considering the following example. The first case is to consider each individual description s and t very good and have the distortion $D_s = 2^{-2R_s}$ and $D_t = 2^{-2R_t}$ respectively. Then Eq. (2.6) can be written as,

$$D_c \geq \frac{D_s D_t}{D_s + D_t - D_s D_t}. \tag{2.8}$$

By assuming some more inequalities, Eq. (2.8) becomes $D_c \geq \frac{min(D_s, D_t)}{2}$, which means the joint decoding is slightly better than the better of the individual description decoding. The second case is to consider the joint decoding description as good as possible so that $D_c \geq 2^{-2(R_s+R_t)}$, which means $\gamma = 1$, therefore, Eq. (2.7) becomes

$$D_s + D_t = 1 + 2^{-2(R_s+R_t)}. \tag{2.9}$$

It is clear from Eq. (2.9) that either D_s or D_t is similar to D_c and the other description has the distortion value 1. The distortion value 1 is obtained by estimating the source from its mean value, which means the description is useless and has no information. The intermediate to the above two example scenarios for the multiple description region is obtained by considering balanced description case, where, $R_s = R_t \gg 1$ and $D_s = D_t \ll 1$. By considering these condition the value of γ in Eq. (2.7) becomes

$$\begin{aligned}\frac{1}{\gamma} &= 1-\left(\sqrt{(1-D_s)(1-D_t)}-\sqrt{D_sD_t-2^{-2(R_s+R_t)}}\right)^2 \\ &= 1-\left((1-D_s)-\sqrt{D_s^2-2^{-4R_s}}\right)^2 \\ &\approx 1-((1-D_s)-D_s)^2 \\ &= 4D_s-4D_s^2 \\ &\approx 4D_s.\end{aligned} \tag{2.10}$$

By substituting the value of γ from Eq. (2.10) into Eq. (2.6) the central distortion value becomes $D_c \geq 2^{-4R_s}(4D_s)^{-1}$. The lower bound of the rate distortion region is obtained by the product of the side and central distortion *i.e.,* $4^{-1}2^{-4R_s}$. Therefore the best decay of central distortion is $D_c \geq 4^{-1}2^{-2R_s}$.

It is clear from the above discussion that it is difficult to utilize $R_s + R_t$ bits together if good descriptions are designed at rate R_s and R_t and transmitted over channel 1 and channel 2 respectively. Similarly, if a good representation of source is done at rate $R_s + R_t$ then it is difficult to split into two useful descriptions. Therefore, the tradeoff in designing any practical MDC scheme is to make the individual description good, but not too similar.

2.5 Multiple Description Coding of Images

The conventional transform-based image coder consists of transform, quantization and entropy coding blocks to remove interpixel, psychovisual and statistical redundancies respectively. The MDC is a source coding method to create multiple descriptions by adding some controlled amount of redundancy among descriptions to protect the source from channel errors. Therefore, the first and important point in any multiple description image coding scheme is to find the stage, where the source is divided into two or more descriptions and redundancy is added easily and effectively. A very simple way of generating multiple descriptions is to create different spatial versions of the image by downsampling followed by individual encoding of each downsampled image by any existing transform-based image coder. The downsampling methods considered in such MDC algorithms include quincunx sampling [60, 61] or polyphase decompositions [6, 62, 63].

Most of the MDC designs are based on integrating multiple descriptions into usual coding modules, such as, the decorrelating transform and quantization [4, 5, 64–75]. Notable examples for modifying the transformation block includes an extension of the JPEG coder as an MDC scheme by using a pairwise correlating transform (PCT) [69]. It was later modified into any number of descriptions [70] and made use of lapped transforms, such as the lapped orthogonal transform (LOT) [71] and the time domain lapped transformed (TDLT) [72]. However, the most commonly used MDC method is modifying the quantization process in a source coder and famously known as MDSQ [5, 65]. Comparison of the wavelet-based image coder with the DCT-based image coder shows that high compression ratio or coding gain is achieved by using the wavelet-based coder [18]. Therefore, many wavelet transform-based MDC schemes are available [5, 66, 73–75] for achieving high coding gain and incorporating extra features, such as scalable decoding. Multiple description image coding approaches can be categorized into following three classes based on different blocks introduced in the conventional image coding model.

1. Multiple Description Coding using Multiple Description Scalar Quantizer.
2. Multiple Description Coding using Pairwise Correlating Transform.
3. Multiple Description Coding using Pre and Post Processing.

2.5.1 Multiple Description Coding using Multiple Description Scalar Quantizer (MDSQ)

The most commonly used approach in creating multiple descriptions is based on modifying the quantization block. Vaishampayan proposed an idea to create multiple descriptions using quantization and is known as MDSQ. The optimal design of fixed rate MDSQ and good index assignment for a memoryless Gaussian source has been studied previously [67]. The optimal design of entropy

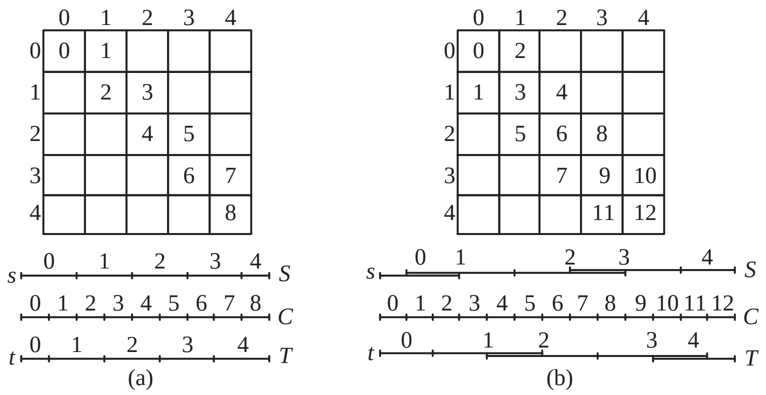


Figure 2.4: Two examples of index assignment matrix and the corresponding central and side quantizers (a) staggered case index assignment (b) modified nested index assignment.

constrained MDSQ is discussed in [76], while the high rate analysis of fixed rate and entropy constrained MDSQs is derived in [77].

An MDSQ consists of two parts: A scalar quantizer that maps a set of random variables $O \in \{o_0, o_1, o_2, ...\}$ to another countable set $C \in \{0, 1, 2, ..., n-1\}$ (commonly known as the central quantizer) and an index assignment matrix that splits the indexes of the central quantizer into two complementary and redundant descriptions, commonly called the side quantizers. The reconstructed quality of the source from the side quantizers is lower than the reconstructed quality from the central quantizer. The relationship of the quantizer bins in the central quantizer to those in the side quantizers are defined by an index assignment matrix, whose row and column indexes correspond to those of the side quantizers, S and T, respectively. The amount of redundancy between the descriptions is controlled by the number of diagonals, f, filled in the index assignment matrix. Two different index assignment matrices and their corresponding central and side quantizers are shown in Figure 2.4.

Figure 2.4(a) shows an example of having staggered index assignment resulting in side quantizers with non-overlapped quantizer bins, while Figure 2.4(b) shows an example of having a modified nested index assignment resulting in side quantizers with overlapped quantization bins. The two descriptions are created by the row and column indices of the index assignment matrix. In either case, the number of diagonals (f) filled in the index assignment matrix defines the maximum side quantizer bin spread, *i.e.,* $g\delta$, where δ is the quantizer bin width of the central quantizer and g is the maximum side quantizer bin spread factor. The value of g for the two cases, the staggered and the modified nested index assignment matrix, is obtained by Eq. (2.11) and Eq. (2.12), respectively

$$g = f, \tag{2.11}$$

$$g = \frac{f^2}{2} - \frac{f}{2} + 1. \tag{2.12}$$

The reconstructed value is the same as the central quantizer reconstruction when both the descriptions are received. On the other hand, the side reconstruction quality depends on the number of diagonals filled in index assignment matrix. In Figure 2.4(a) only 9 cells of the index assignment matrix are filled. The redundancy between the descriptions depends on the unfilled cells. The highest possible redundancy between the descriptions is achieved by filling only the main diagonal of the index assignment matrix, resulting in similar central and side decoding quality. The side decoding quality is lowest if all the cells of the index assignment matrix are occupied, resulting in no redundancy between the descriptions.

The very first multiple description image coding based on MDSQ is proposed by Vaishampayan [78], in which the MDSQ is applied to the DCT coefficients of the JPEG coder. After MDSQ, two descriptions are then entropy coded separately and transmitted through different channels. A wavelet and MDSQ-based multiple description image coding is proposed by Servetto in [5]. In this coder, MDSQ is applied to the wavelet coefficients and better redundancy allocation is achieved by using different index assignment matrices for different subbands. The SPIHT algorithm then encodes the independently created descriptions. Conventional MDC schemes focus on generating two descriptions having balanced rate distortion performance. Most of the methods on generating more than two descriptions are based on subsampling of the source content [5, 66, 79]. Examples include wavelet zero tree-based subsampled packetization of the two descriptions generated from a single MDSQ [5] and grouping together of different wavelet trees using the SPIHT algorithm [66]. However, the main problem of these schemes is the severe effect on the reconstructional quality of the content on joint decoding when any of the description is unavailable, as each description carries the information of certain coefficients of the image. In Chapter 3, a new approach for generating more than two descriptions is proposed, with each description containing information from all coefficients and yielding descriptions with balanced and unbalanced rate distortion performance.

Different MDSQ-based methods are proposed to incorporate quality scalability with in the multiple description image coding framework [53, 73, 80, 81]. In [73], a layered tree-based multiple description coding scheme is presented. In [53], a scalable multiple description coding (SMDC) scheme based on embedded MDSQ (EMDSQ) and quad tree type coding is discussed in detail. In EMDSQ, a set of side quantizers that generates different number of descriptions is derived from an embedded central quantizer. The index assignment matrix considered in EMDSQ results in side quantizers with non-overlapped cells and only the balanced joint decoding is possible. The design problem of EMDSQ is discussed in detail in [54, 82, 83]. In contrast to EMDSQ, the MDSQ-SR presented in Chapter 4 considers different index assignment matrices resulting into non-overlapped and overlapped side quantizer bins to incorporate different amount of redundancy between the descriptions at the base layer. MDSQ-SR not only facilitates the user to incorporate different amount of redundancy among the descriptions depending on the number of diagonals filled in the index assignment matrix of the base layer but also supports the joint decoding in balanced and unbalanced fashion.

2.5.2 Multiple Description Coding using Pairwise Correlating Transform (PCT)

In the previous subsection MDC scheme is discussed, in which multiple descriptions are generated by the index assignment matrix after the quantization block of conventional image coder. In this subsection, another MDC method is discussed for images, in which redundancy is added immediately after the transformation block of the conventional image coder and is called PCT. In conventional image coding, transformation is used to decorrelate the input image. However, in PCT-based MDC schemes, controlled amount of correlation is introduced in the transformed coefficients. The main objective of the PCT is to estimate the lost coefficients from the received ones.

Figure 2.5 shows a general framework of the multiple description image coding

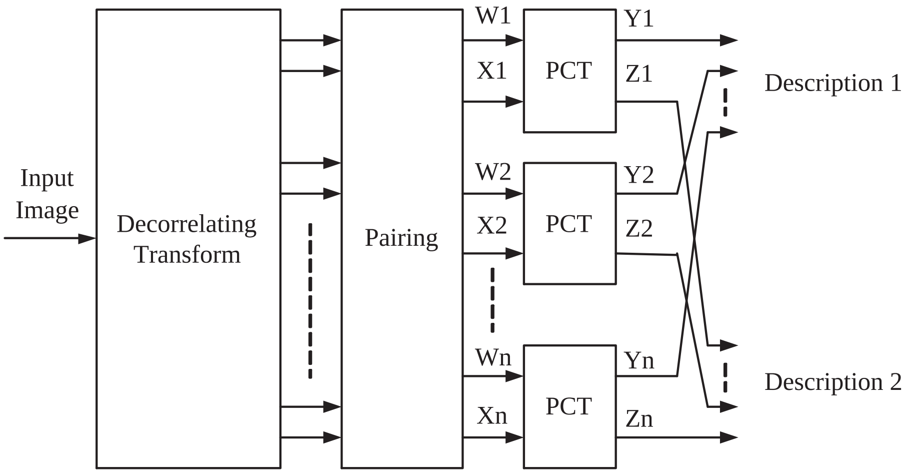


Figure 2.5: MDC using PCT for two descriptions.

based on PCT for $N = 2$ descriptions. Firstly, the input image is decorrelated by decorrelating transform *i.e.,* DCT or DWT. For two-description case, the transformed coefficients are arranged according to their variance and grouped into pairs. The pairing of the coefficients is performed in such a manner to minimize the estimation error when the missing PCT coefficients are estimated from the available coefficients. In [69], the coefficients with a large and small value are paired together. The PCT block take the uncorrelated coefficient pairs as an input and introduce certain amount of redundancy between the coefficients to generate the correlated coefficient pairs as an output. After PCT, two descriptions are generated by sending one coefficient of pair to description 1 and the other coefficient to description 2 as shown in Figure 2.5. On the decoder side, when both the descriptions are received the coefficients can be recovered exactly by applying inverse PCT. On side decoding, inverse PCT is applied after estimating the missing coefficient from the received one. The PCT used in multiple description image coding can be orthogonal [84] or non-orthogonal [85, 86].

The very first MDC scheme based on orthogonal pairwise correlating transform is proposed in [84]. Let $\mathcal{W}$ and $\mathcal{X}$ be the two Gaussian random variables with variances ${\sigma_{\mathcal{W}}}^2$ and ${\sigma_{\mathcal{X}}}^2$ respectively. Let $\mathcal{Y}$ and $\mathcal{Z}$ be the output of the PCT having variances ${\sigma_{\mathcal{Y}}}^2$ and ${\sigma_{\mathcal{Z}}}^2$ respectively. The variables $\mathcal{Y}$ and $\mathcal{Z}$ are related to $\mathcal{W}$ and $\mathcal{X}$ by the transformation matrix $\mathcal{T}$ as $[\mathcal{W}\mathcal{X}] = \mathcal{T}[\mathcal{Y}\mathcal{Z}]$. The transformation matrix $\mathcal{T}$ adds the correlation between the descriptions, therefore controlling the redundancy as well. The transformation matrix used for orthogonal PCT in [84] is

$$\mathcal{T} = \begin{bmatrix} cos\phi & sin\phi \\ -sin\phi & cos\phi \end{bmatrix}, \tag{2.13}$$

where ϕ controls the redundancy between the description. The non-orthogonal PCT is also developed for multiple description coding. Pair selection criteria, redundancy allocation and redundancy rate distortion analysis of the orthogonal and non-orthogonal pairwise correlating transform are briefly explained in [64]. The optimal transform matrix that generates equal side distortion and balanced rate description is proposed in [85] and is written as,

$$\mathcal{T} = \begin{bmatrix} \sqrt{\frac{cot\phi}{2}} & \sqrt{\frac{tan\phi}{2}} \\ -\sqrt{\frac{cot\phi}{2}} & \sqrt{\frac{tan\phi}{2}} \end{bmatrix}. \tag{2.14}$$

It is shown in [64] that the non-orthogonal PCT outperforms the orthogonal PCT in terms of redundancy, rate and distortion. Both the orthogonal and non-orthogonal PCT gives the same result at maximum and minimum level of redundancy allocation. The maximum redundancy is achieved in orthogonal PCT when $\phi = \frac{\pi}{4}$. It is also observed from Eq. (2.13) and Eq. (2.14) that both the orthogonal and non-orthogonal PCT are same when $\phi = \frac{\pi}{4}$. Comparison of PCT-based MDC with MDSQ-based MDC shows that PCT gives better performance at low redundancies than MDSQ [85]. The problem presented in [64,69] are only for two descriptions, which is generalized to any number of descriptions N in [4,70]. A Cascade structure is used to generalize the PCT-based MDC scheme to any arbitrary number of descriptions N. It is shown in [68] that the DWT-based PCT method of MDC outperforms the DCT-based PCT method both objectively and subjectively.

2.5.3 Multiple Description Coding using Pre and Post Processing

The MDC schemes discussed in Section 2.5.1 and Section 2.5.2 are stand-alone techniques and are not compatible with the standard image coders like JPEG or JPEG2000. Different MDC methods have been proposed that use the standard

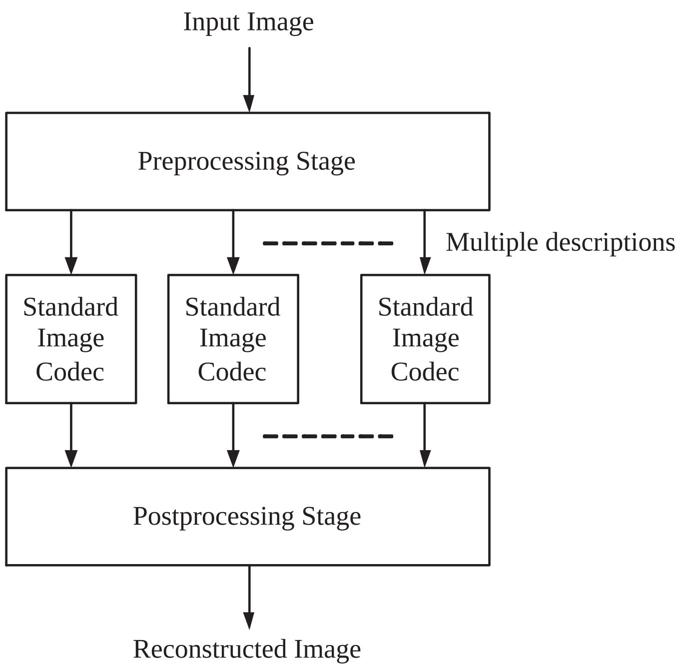


Figure 2.6: MDC using pre and post processing.

coders to encode each description after preprocessing stage [6, 60–63]. Multiple descriptions are generated by downsampling the input image followed by individual encoding of each downsampled image by an existing coder. On the decoder side, each description is first decoded by the standard decoder and then combined to generate the output image. Figure 2.6 shows a framework of preprocessing-based MDC scheme for images. The preprocessing stage consists of two steps: one is to create multiple descriptions and the other is to insert redundancy within each description.

Multiple descriptions in preprocessing stage is created by either polyphase down sampler as in [6, 62, 63] or by using the Quincunx sampling as in [60, 61] to exploit the spatial correlation present in the images. Such type of subsampling leads to create any arbitrary number of descriptions according to the channel conditions. In case of any description lost at the decoder, the missing pixels are interpolated from the received neighbour pixels. The interpixel redundancy within each subsampled image is decreased by subsampling, resulting in increase in data rate, therefore adding some redundancy in each description. The amount of redundancy is also increased by adding other polyphase components in each description in addition to the main polyphase component of that description [87]. In such schemes, each description contains one main polyphase component that

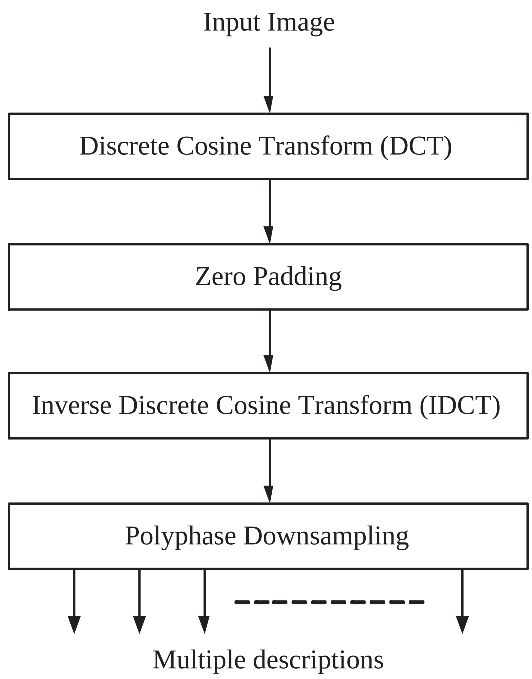


Figure 2.7: Blocks involved in preprocessing stage.

is coded at higher data rate and some additional polyphase components that are encoded at much lower data rate than the main polyphase component. In case of main polyphase component loss, the available low data rate version of the same polyphase component from the other description can be used. The rate allocation for each polyphase component and redundant copies of polyphase component in each description is well studied in [87].

The redundancy in each description can also be added at the preprocessing stage before the spatial subsampling. In such type of redundancy insertion, the resolution of the input image $X_1 \times Y_1$ is increased to the resolution $X_2 \times Y_2$, where $X_1 < X_2 < 2X_1$ and $Y_1 < Y_2 < 2Y_1$. The procedure of adding redundancy in preprocessing stage is shown in Figure 2.7. Firstly, the input image of size $X_1 \times Y_1$ is transformed by the DCT and then zeros are padded to increase the resolution of the input image. The number of padded zeros controls the redundancy within each description. The amount of zeros added should be less than the number of zeros required to double the original resolution. Otherwise, it is same as sending the same image twice. After zero padding, inverse DCT is performed to reconstruct the image at higher resolution. Any arbitrary number of descriptions is then generated by polyphase decomposition.

After creating multiple descriptions, each description is coded by using standard image coder like JPEG or JPEG2000. On the decoder side, the standard decoder first decodes each description. In case of description loss, estimation or prediction is performed for missing descriptions. Interpolation [6, 61] and image fusion [60] techniques are used to reconstruct the image at higher quality when more than one descriptions are received at the decoder.

2.5.4 Comparison of Multiple Description Image Coding

The multiple description image coding methods discussed so far are based on modifying any of the block or by adding some extra blocks in conventional image coding model. There are other methods available in the literature that creates the multiple descriptions using some existing image coding algorithms. A SPIHT-based multiple description image coder to create arbitrary number of descriptions is proposed in [66]. The N descriptions are generated by grouping spatially dispersed wavelet coefficient trees, which are then encoded with SPIHT. In each description, one of the spatially dispersed wavelet coefficient tree group is coded at higher rate by considering higher number of refinement passes. The other tree groups are coded with much lower rate to add redundancy in that description. The amount of redundancy is controlled by varying the number of refinement passes of each tree. If any of the description is lost, the missing trees are obtained from the best available quality copy of the missing tree present in other descriptions.

Similar kind of MDC method based on JPEG2000 is proposed by Tillo [88]. In this method, two JPEG2000 bitstreams are generated at two different rates. As the data rate allocation in JPEG2000 is based on code blocks truncation, therefore the code block in two bitstreams encoded at different data rates have different code block truncation points. Two balanced descriptions are created by considering different code blocks from different JPEG2000 bitstreams.

The choice of the multiple description image coding depends on the complexity and adaptability of the scheme. Pre and post processing-based MDC approaches are very useful in terms of adaptability as these schemes use standard image codecs. On the other hand, for the redundancy allocation algorithms, extra pre and post processing stage increase the complexity of these schemes. MDSQ-based MDC provides the easiest way to insert the redundancy between the descriptions. The amount of redundancy in MDSQ is controlled by varying the number of diagonals filled in the index assignment matrix. However, the redundancy insertion in PCT and pre and post processing based methods are more complicated than MDSQ based method. In PCT-based MDC, an extra PCT block is required to insert redundancy while in preprocessing-based MDC an upsampling of the original image is required to add redundancy between the descriptions. Similarly, on the decoder side, estimation of the missing description is required for the PCT-based methods and interpolation and fusion techniques are required for pre and post processing based methods. From the state-of-the-art analysis of multiple description image coding, it is evident that MDSQ-based MDC scheme gives better performance and are most feasible to insert different amount of redundancies between the descriptions. Therefore, the solution provided in this thesis for generating any number of scalable descriptions is based on MDSQ.

2.6 Multiple Description Coding of Video

In most of the video coding standards, temporal correlation is exploited by motion compensated prediction, which splits the video into prediction error and motion vectors. So for complete multiple description video coding (MDVC) system, both the prediction error and motion vectors are coded in multiple streams. Multiple descriptions of the prediction error can be created similarly as in multiple description image coding by using PCT [64, 89], MDSQ [90, 91], and by spatial subsampling of the transformed prediction error coefficients [52]. Similarly, multiple descriptions of the motion vector information can be generated either by duplication or by subsampling of the motion vector information. Another way of creating multiple descriptions is based on subsampling in temporal domain [9, 92]. In this section, different MDVC methods and problems related to them are discussed. MDVC methods can be categorized into following two categories.

1. Predictive Multiple Description Video Coding.

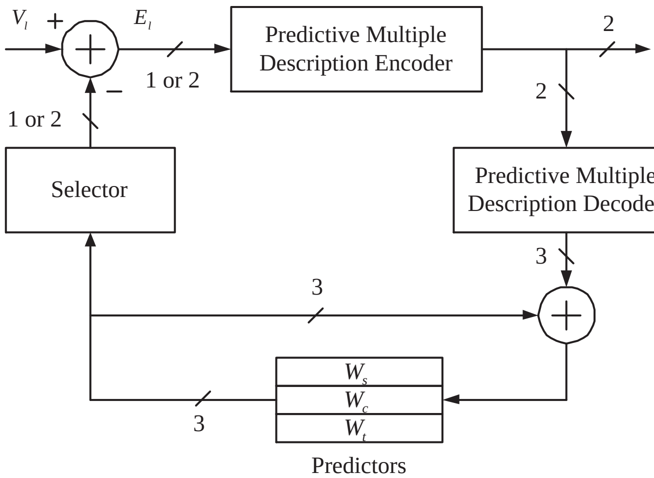


Figure 2.8: Block diagram of generic predictive multiple description video encoder.

2. Scalable Multiple Description Video Coding.

2.6.1 Predictive Multiple Description Video Coding

Motion compensation and prediction is a fundamental component of any video coding system as it exploits the temporal redundancy between the frames. In single description video coding environment, the encoder predicts the information by assuming the state at the decoder. Therefore, for the perfect reconstruction of video, both the encoder and decoder maintain an identical state. However, MDC schemes are designed on the principle that some of the information may not always be available at the decoder due to loss during the transmission. The mismatch between the encoder and decoder can occur whenever an encoder predicts the signal, which is not available at the decoder. Therefore, mismatch control is a major concern in predictive MDVC. The mismatch problem can be controlled by using extra prediction loops or by duplicating the prediction error in each description.

Figure 2.8 shows a block diagram of generic predictive multiple description video encoder for two descriptions. In Figure 2.8 each arrow may contain one, two, or three signals. Different predictive MDVC methods are proposed based on different number of prediction loops, by omitting or adding different number of signals in each arrow of the generic encoder. The generic encoder works as follows. Depending on the implementation, the encoder stores three previous frames, W_s, W_t, and W_c reconstructed from individual descriptions s and t and by joint description decoding respectively. The motion compensated difference E_l from the central predictor that uses W_c, are coded using any multiple description encoder to generate two descriptions. The encoder also has multiple description decoders that reconstruct the error signal when single or both descriptions are received. There is no mismatch if both the descriptions are received at the decoder. But if the single description is received at the decoder, some information is missing leading to the mismatch condition. To avoid the mismatch, the side prediction error *i.e.,* $V - W_s - W_c$ and $V - W_t - W_c$ can be embedded in description s and t respectively. The side description error controls the amount of redundancy between the descriptions and mismatch between the encoder and decoder. The predictive MDVC can be categorized into three categories and are discussed below, depending on the number of prediction loops in the encoder that controls the mismatch completely or partially.

2.6.1.1 Single Predictor with Mismatch

In this type of predictive multiple description encoders, the same predictor is used as in single description predictive encoders. The prediction error is minimized by the single predictor and no additional redundancy is added in generating multiple descriptions during the prediction process except the redundancy added by the multiple description method used to create two descriptions of the prediction error [7,93,94]. The prediction can only be produced correctly when both the descriptions are available at the decoder. This type of coders results in a mismatch when any of the description is not available at the decoder.

The very first predictive MDVC scheme with no mismatch control is based on PCT [95]. Single description predictor is used to generate the prediction error. The DCT is applied to the prediction error followed by the PCT. Two descriptions are created by sending one member of each pair in each description. The motion vector information is repeated in both descriptions. As there is no mismatch coding involved, therefore predictive intra frames are inserted in each description repeatedly to clear up the mismatch.

Instead of using PCT to create multiple descriptions, a very simple method based on alteration and duplication with optimal rate-distortion strategy is proposed in [93]. A single bitstream generated from H.263 encoder is used to create two descriptions. The motion vector information and low frequency coefficients are duplicated in each description while the high frequency coefficients are alternated between the descriptions. The number of duplicated coefficients controls the amount of redundancy between the descriptions. The amount of redundancy due to duplication of motion vector is reduced in [94] by splitting the motion vectors and DCT coefficients of adjacent blocks into two description using quincunx sampling. In [93], only low frequency coefficients are duplicated in each description but in [96] one description is composed of all coefficients and the other description only contains the low frequency coefficients resulting into unbalanced descriptions. Similarly in [97], the splitting method of creating multiple description is extended that allows any frequency coefficient to be duplicated in any description.

2.6.1.2 Multiple Predictors with No Mismatch

In this type of predictive multiple description encoding, both the encoder and decoder creates the same prediction signal resulting in no mismatch. There are two different ways to avoid the mismatch. First one is to use independent predictors for each description [7, 98–100] and other is to use single predictor that uses the information from both the descriptions [95].

In [99, 101] video redundancy coding (VRC) algorithm is proposed that divides the input video into two subsets by temporal subsampling. Each Temporally subsampled sequence is then encoded by using independent single predictive video encoder. As each description has its own predictor, therefore there is no mismatch even if a single description is available at the decoder. In case of single description decoding, the missing frames are estimated from the received frames. Temporal subsampling decreases the correlation among the frames, therefore, introduce the redundancy between the descriptions. The other source of redundancy in VRC is due to the synchronization frames coded in each description. Another subsampled based error resilient video coding, which has used the concept of multiple states is proposed in [102]. In this scheme, video sequence is split into even and odd frames and then encoded with its own predictor and state information. The video stream is recovered at half of the original frame rate if a single description is available. In contrast to VRC scheme, there is no requirement of synchronization frame in multiple state coding because the state recovery can be done by using the information from both streams.

Instead of using temporal subsampling to create multiple descriptions, polyphase downsampler is used in independent flow multiple description video coder (IF-MDVC) to generate multiple descriptions [7]. The redundancy between the description is introduced by the preprocessing stage that used DCT and zero padding as explained in Section 2.5.3. To avoid the mismatch, motion compensation prediction is performed on each flow separately. In case of description loss during transmission, the missing samples are obtained by interpolating the received samples.

2.6.1.3 Single Predictor With None or Partial Mismatch

Mismatch problem in predictive multiple description video encoders can be avoided completely or partially by including mismatch coding using the different methods discussed in Section 2.6.1.1. A very simple example to partially control the mismatch is to use the intra coded frames in each description at regular intervals to reset the mismatch errors. Another approach to control mismatch is to use three prediction loops as shown in the diagram of generic multiple description predictive encoder.

In [95], two different algorithms are proposed that solves the mismatch problem partially and completely by using three prediction loops at the encoder. In both the algorithms, motion vector and header information is duplicated in each description. Intra and motion compensated prediction error frames are coded into multiple descriptions by using PCT. In algorithm 1, mismatch is avoided completely by embedding the side prediction error in each description. Similarly, in second algorithm mismatch is controlled partially by embedding partial amount of side loop prediction error. Therefore, results into reduction in redundancy between the descriptions.

2.6.2 Scalable Multiple Description Video Coding

The MDVC schemes discussed in the previous section are based on non-scalable predictive video coding framework. The advantage of those schemes is that one can use the existing standardized coding blocks in MDC framework, which are designed for high coding gain. However, high coding gain is not always the only requirement especially in case of heterogeneous bandwidth requirements. Scalable coding addresses the problem of heterogeneity but is prone to errors. On the other hand, MDC is a source coding scheme, which is robust against transmission losses. Therefore, different schemes that combine the scalable and multiple description coding are proposed to provide both the scalability and reliability. Scalable multiple description video coding (SMDVC) schemes can be categorized into two categories.

2.6.2.1 3D Transform based Scalable Multiple Description Video Coding

In this section, open loop architecture of SMDVC based on 3D transform is discussed. The open loop architecture uses the MCTF and multi resolution spatial decomposition that provides the scalability and gives better performance than the close loop scalable video coding [103]. A general framework of the 3D transform based MDVC is shown in Figure 2.9. Firstly, the temporal correlation among different frames is removed by applying MCTF. The MCTF block converts the input video into low and high frequency frames and a set of motion vectors. The low and high frequency frames are then be decomposed by 2D spatial transform to remove the spatial redundancy and hence complete the 3D saptio-temporal decomposition. Different methods are available in the literature on generating multiple descriptions from the spatio-temporal decomposed coefficients.

In [10], Bajic and Woods proposed a MDC method for images and video based on data partitioning. Multiple descriptions are generated for images by partitioning

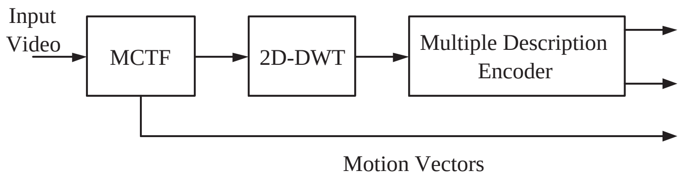


Figure 2.9: 3D transform based multiple description video coding framework.

the wavelet transformed coefficients in such a way that each description contains the coefficients that are maximally separated from each other. By using data partitioning no redundancy is added in each description and errors can be concealed from the received descriptions. In case of video, the 3D spatio-temporal coefficients and the motion vectors are partitioned into different descriptions in a similar way as in images. In [11], multiple descriptions are generated from the spatio-temporal coefficients by repeating the low frequency frames and motion vector information in each description and high frequency frames are divided between the two descriptions. The redundancy is controlled by the amount of duplicated information in each description. In case of single description decoding, the missing frames are estimated from the motion vector information, therefore the quality of single description decoding is low for sequences having high motion.

Another 3D wavelet transform based scheme that generates flexible number of descriptions is proposed in [104]. The main concept of this scheme is to encode each spatio-temporal decomposed code block to any given number of descriptions. Each code block is encoded at high data rate for one description and at low data rate for all other descriptions. Different code blocks are then mixed from high data rate and low data rate encoding to generate a description. A rate scalable coding scheme (EBCOT) is adopted to encode each code block. If all the descriptions are available at the decoder, the code blocks encoded at high data rate from each description are selected for decoding. On single description decoding, still an acceptable but low decoding quality is possible from the low data rate code blocks. The redundancy is controlled by the rate-distortion selection of each code block. In [12], embedded MDSQ proposed in [54, 82, 83] is used to generate scalable descriptions from the spatio-temporal decomposed coefficients. The channel aware rate allocation algorithm is also adopted in this scheme to further improve the reliability of the video delivery under packet and bursty losses. Inspired from the results of MDSQ-SR in SMDC for images, a 3D transform based framework for scalable multiple description video coding is proposed in Chapter 6 that uses the concept of MDSQ-SR. The scalable descriptions of the video generated from several MDSQ-SR are capable of joint decoding at different quality, resolution and frame rate in balanced and unbalanced manner. Two schemes that reduces the texture and motion vector redundancies are also proposed in Chapter 6.

2.6.2.2 Standard Compatible Scalable Multiple Description Video Coding

In the previous section, 3D transform based SMDVC schemes are discussed, which are not compatible with the standard video coders. In this section, the SMDVC schemes that are compatible with the standard codecs are presented. There are two different ways to create scalable descriptions from single scalable video coder. One is to create multiple descriptions as done in VRC and then make them scalable [9,13,14] and the other is by mapping the single scalable description into multiple descriptions [8,15–17].

In [16], scalable multiple descriptions are generated by duplicating the base layer of the H.264/SVC bitstream in each description and half of the texture and motion vector information from H.264/SVC enhancement layers is embedded in each description. Another H.264/SVC compatible SMDVC scheme is proposed that uses the same concept as in [104], where low and high data rate bitstreams are extracted by considering different video segments from the H.264/SVC bitstream. A very simple example to create H.264/SVC compatible description is to extract even frames at full quality and only the base layer for odd frames and vice versa for the other description. In [17], five different methods are proposed that generates two descriptions from H.264/SVC bitstream by considering different redundancy levels and rate-distortion performance of each description.

The second type of standard compatible SMDVC schemes are based on temporal subsampling. In [13], two descriptions are created by considering even and odd frames and then each subsampled sequence is independently encoded by using H.264/SVC codec. In case of description loss, the missing frames are replaced by interpolating the frames from the received description. The interpolation technique used in [13] is simple averaging, therefore gives a ghosting effect on side decoding for high motion sequences. The ghosting effect can be reduced by sending extra side information in each description. The temporal subsampled-based SMDVC scheme with residual coding is presented in [9]. The residual side information is generated by subtracting the adjacent frame from the interpolated frame. The residual information from the other description is then embedded in each description. The side information increases the redundancy between the descriptions and also improves the side decoding quality as compared to the method in [13]. Instead of using simple frame averaging, ghosting effect can be reduced by motion based interpolation algorithm resulting in low redundancy as well.

2.7 Conclusions

In this chapter, various MDC methods for images and video are discussed and broadly categorized into quantization, transform and pre and post processing based categories. Most of the MDVC methods are based on single description video coding. The subsampled based standard compatible schemes results in ghosting effects in case of description loss. On the other hand, the mapping of one scalable bitstream into multiple description results into high redundancy and low joint quality gain. From the state-of-the-art analysis, it is also evident that quantization based MDC schemes are most feasible to insert different amount of redundancies between the descriptions. Therefore, in this thesis, we preferred quantization based MDC approach to generate any number of scalable descriptions that can be decoded individually and jointly at any quality, resolution and frame rate in balanced and unbalanced manner.

Chapter 3

Multi-Channel Unbalanced Description Coding

In the previous chapter, the MDC schemes discussed are mainly based on two balanced descriptions. In this chapter, a new MDC method is presented that generates and jointly decodes any number of rate-distortion wise balanced and unbalanced descriptions using several MDSQs. The proposed method starts with an initial MDSQ design that generates two descriptions and different MDSQs are obtained by refining the central quantizer, resulting in a hierarchy of MDSQs. The joint decoding distortion constraints for different combination of descriptions from several MDSQs are formulated with the objective to improve the distortion as the number of descriptions jointly decoded is increased. For meeting these constraints, the design conditions for several MDSQs are proposed. The rest of the chapter is organized as follows. Section 3.1 provides an overview of the MDC schemes having more than two descriptions. The design constraints and conditions of the MDC schemes using several MDSQs is presented in Section 3.2. Section 3.3 demonstrates how multi-channel unbalanced description coding scheme is extended to achieve robust quality scalability within the proposed MDC framework by successive side quantizer bin merging (SSQBM) to design a hierarchical tree of MDSQs and choosing descriptions from the specified nodes of the tree. The performance of the proposed method is evaluated under lossless and lossy channel conditions in Section 3.4.

3.1 Background

Conventional MDC schemes focus on generating two descriptions having balanced rate-distortion performance. Most of the methods on generating more than two descriptions are based on sub-sampling of the source content by relaxing the property that every single coefficient information is used in each description [5, 66, 79]. Examples include wavelet zero tree-based sub-sampled packetization of the two descriptions generated from a single MDSQ [5] and grouping together of different wavelet trees using the SPIHT algorithm [66]. The main problem of these schemes is the severe effect on the reconstructional quality of the content on joint decoding when any of the description is unavailable, as each description carries the information of certain coefficients of the image. In another method [105], a generalized multiple description scheme is proposed by dividing the side quantizer into equal length blocks corresponding to the required number of descriptions and generating balanced descriptions by sequentially distributing the cyclic permutations of the blocks in each description. In this chapter, a new approach for generating more than two descriptions is proposed, with each description containing information from all coefficients and yielding descriptions with unbalanced rate-distortion performance, leading to a multi-channel unbalanced description coding (MUDC) scheme. The proposed scheme generates an arbitrary number of descriptions by using several hierarchically related MDSQs and provides the user an option to select descriptions from different MDSQs having different rate-distortion properties leading to balanced and unbalanced joint decoding. This scheme hierarchically refines the central quantizer and exploits the resulting hierarchical quantizers on achieving multi casting-based robust quality scalability. The MUDC method not only provides the quality scalability but also provides the flexibility to add and remove overall redundancy in terms of number of descriptions.

3.2 Multiple Descriptions using Several MDSQs

As discussed in Section 2.5.1, an MDSQ generates two side quantizers resulting in two redundant descriptions. To generate any arbitrary number of descriptions N, J number of MDSQs are considered, where $J \geq \frac{N}{2}$, as shown in Figure 3.1.

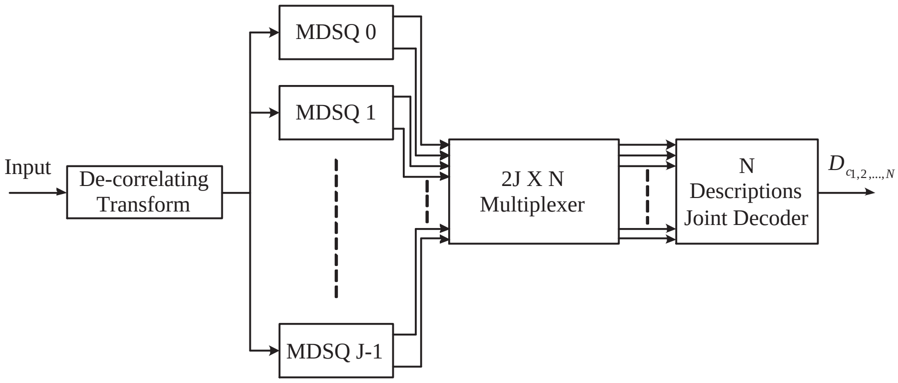


Figure 3.1: MDC scheme for N descriptions using J MDSQs.

The output from the decorrelating transform is passed to at least J number of MDSQs to generate $2J$ descriptions. Let s_m and t_m be the descriptions generated from the side quantizers, S_m and T_m, respectively derived from the central quantizer C_m, where $m = 0, \cdots, J-1$ is the MDSQ index. The three quantizers, C_m, S_m and T_m, is depicted as a three node tree with C_m representing the root node and the other two representing leaf nodes. It is considered that each C_m in J number of MDSQs have unique rate-distortion properties, generating two balanced descriptions according to the rate-distortion conditions in Eq. (2.3). It is also assumed that the MDSQs are arranged and indexed in descending distortion, with m=0 corresponding to the highest distortion and $m = J-1$ corresponding to the lowest distortion. This allows the side descriptions resulting from any side quantizer from different MDSQs to form a set of unbalanced descriptions. A few examples (for unbalanced $N = 2$, $N = 3$ and $N = J-1$ descriptions) are shown in Figure 3.2. The joint decoding distortion constraints for different combination of descriptions from several MDSQs need to be formulated with the objective to improve the distortion as the number of joint decoding descriptions is increased. In next section the distortion constraints for the MUDC scheme are formulated.

3.2.1 Distortion Constraints

The MDSQ-based MUDC problem is formulated by considering the decoding distortion constraints as follows. Let us consider two MDSQs with indexes, $m = j$ and $m = k$, where $k > j$. The distortion in corresponding single description

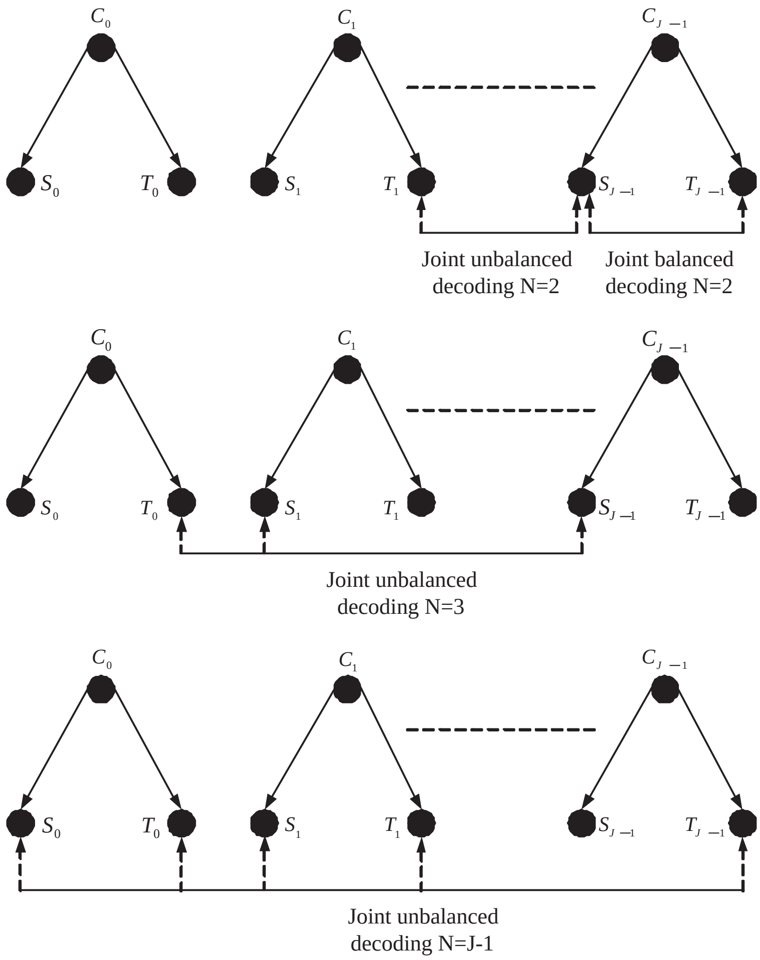


Figure 3.2: Tree structure for J MDSQs and examples of joint decoding for balanced $N = 2$ and unbalanced $N = 2$, $N = 3$ and $N = J - 1$ descriptions.

decoding is related as

$$D_{s_k} < D_{s_j}, \tag{3.1}$$
$$D_{t_k} < D_{t_j}, \tag{3.2}$$
$$R_{s_k} > R_{s_j}, \tag{3.3}$$
$$R_{t_k} > R_{t_j}, \tag{3.4}$$

and the distortion of joint decoding of two balanced descriptions as

$$D_{c_k} < D_{c_j}, \tag{3.5}$$
$$R_{c_k} > R_{c_j}. \tag{3.6}$$

Secondly, the joint decoding of side descriptions form MDSQ j and k are considered, *i.e.,* unbalanced descriptions. The distortion constraints of the joint unbalanced description decoding is set as

$$D_{c_k} \leq \quad D_{c_{j,k}} \quad \leq D_{c_j}, \tag{3.7}$$

$$D_{c_{j,k}} \quad \leq \quad D_{s_k}, D_{t_k}, \tag{3.8}$$

where $D_{c_{j,k}}$ denotes the joint decoding distortion of unbalanced descriptions. The joint unbalanced decoding distortion has to be better than joint decoding distortion of the descriptions from the MDSQ j and single description decoding of side descriptions s_k and t_k corresponding to the lowest distortion MDSQ.

The distortion of a description created from any MDSQ is related to that of the other depending on the factor g, defined in Eq. (2.11) and Eq. (2.12). For the MUDC scheme, it is assumed that the quantization bin widths of the central quantizers in each MDSQ are related by a scaling factor a^m, resulting in a bin size of $\frac{\delta}{a^m}$, where δ is the bin size of the central quantizer of the first MDSQ. To satisfy the distortion constraints in Eq. (3.1) - Eq. (3.8) for MUDC, the design conditions on the values of a and g for individual and joint decoding scenarios from different MDSQs need to be established.

3.2.2 Design Conditions

For the central quantizer of the MDSQ m with a bin size $\frac{\delta}{a^m}$, the bin size for any of its side quantizers can be written as $\frac{g_m \delta}{a^m}$, where g_m is the maximum side quantizer bin spread depending on the index assignment matrix as in Eq. (2.11) and Eq. (2.12). As shown in Figure 3.3, let y_m and z_m be the minimum and maximum values of any quantizer bin of the central quantizer C_m, respectively. The relationship between two consecutive central quantizers can be written as

$$y_m = y_{m-1} + \frac{i_m \delta}{a^m}, \tag{3.9}$$

where $i_m \in \{0, \cdots, a-1\}$. Now considering all MDSQs from $0, \cdots, j-1$, we can write

$$y_m = y_0 + I_{m,0} \frac{\delta}{a^m}, \tag{3.10}$$

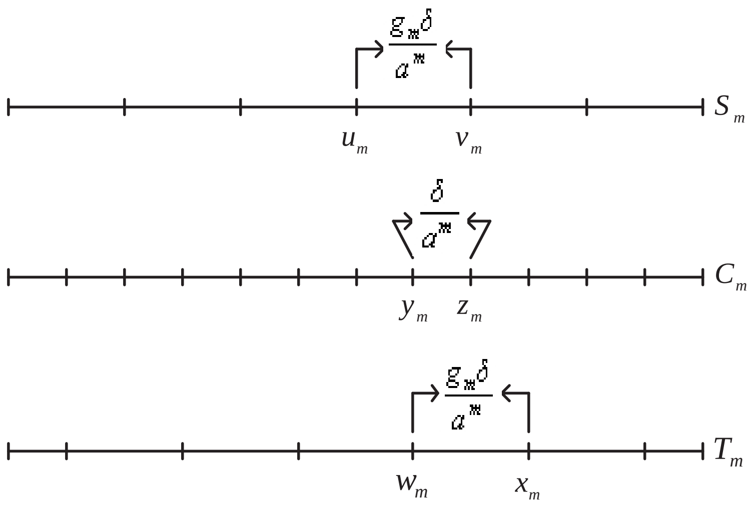


Figure 3.3: Central and side quantizers of an MDSQ.

where $I_{m,0} = \sum_{\alpha=1}^{m} a^{m-\alpha} i_\alpha$.

Let u_m and v_m be the minimum and maximum values of the side quantizer S_m. Similarly, w_m and x_m be the corresponding values of the side quantizer T_m. For any MDSQ, m, the minimum and maximum values of the side quantizer bin is related to the minimum and maximum value of the central quantizer bin as

$$\begin{aligned} u_m &= y_m + (1-h_m)(1-g_m)\frac{\delta}{a^m}, \\ &= y_0 + I_{m,0}\frac{\delta}{a^m} + (1-h_m)(1-g_m)\frac{\delta}{a^m}, \\ &= y_0 + (I_{m,0} + (1-h_m)(1-g_m))\frac{\delta}{a^m}, \end{aligned} \tag{3.11}$$

where h_m is the lead or lag factor of the two side quantizers, *i.e.*, $h_m \in \{0, 1\}$, representing a lag or a lead respectively. Then considering the side quantizer bin spread, $\frac{g_m\delta}{a^m}$, the value of v_m can be written as

$$v_m = y_0 + (I_{m,0} + (1-h_m)(1-g_m) + g_m)\frac{\delta}{a^m}. \tag{3.12}$$

Similarly, for the other side quantizer T_m, the maximum and minimum value of any side quantizer bin is written as

$$w_m = y_0 + (I_{m,0} + h_m(1-g_m))\frac{\delta}{a^m}, \tag{3.13}$$

$$x_m = y_0 + (I_{m,0} + h_m(1-g_m) + g_m)\frac{\delta}{a^m}. \tag{3.14}$$

The implicit central quantizer bin width for the joint decoding of the side quantizers from the same MDSQ (in case of balanced description) and from different MDSQs (in case of unbalanced descriptions) are required.

3.2.2.1 Implicit Central Quantizer Bin Width of the Descriptions from Same MDSQ

Now the bin size Δ_{s_m,t_m} of the implicit central quantizer bin width for the joint decoding of side quantizers from the same MDSQ, m, is formulated as follows:

$$\begin{aligned}
\Delta_{s_m,t_m} &= \min(v_m, x_m) - \max(u_m, w_m), \\
&= \min(u_m + \frac{g_m\delta}{a^m}, w_m + \frac{g_m\delta}{a^m}) - \max(u_m, w_m), \\
&= \frac{g_m\delta}{a^m} + \min(u_m, w_m) - \max(u_m, w_m), \\
&= \frac{g_m\delta}{a^m} - |u_m - w_m|, \\
&= \frac{g_m\delta}{a^m} - \left|\frac{\delta}{a^m}(2h_m - 1)(g_m - 1)\right|. \qquad (3.15)
\end{aligned}$$

By substituting the value of $h_m = 0$ or $h_m = 1$ in Eq. (3.15), the central quantizer implicit bin width, Δ_{s_m,t_m}, becomes

$$\begin{aligned}
\Delta_{s_m,t_m} &= \frac{g_m\delta}{a^m} - (g_m - 1)\frac{\delta}{a^m}, \\
&= \frac{\delta}{a^m}. \qquad (3.16)
\end{aligned}$$

3.2.2.2 Implicit Central Quantizer Bin Width of the Descriptions from Different MDSQs

Now the joint decoding of unbalanced descriptions is considered from different MDSQs, $m = j$ and $m = k$, where $k > j$. That means joint decoding of two side descriptions coming from two different MDSQs. Considering the four descriptions, s_j, s_k, t_j and t_k, there are four possible ways of choosing two unbalanced

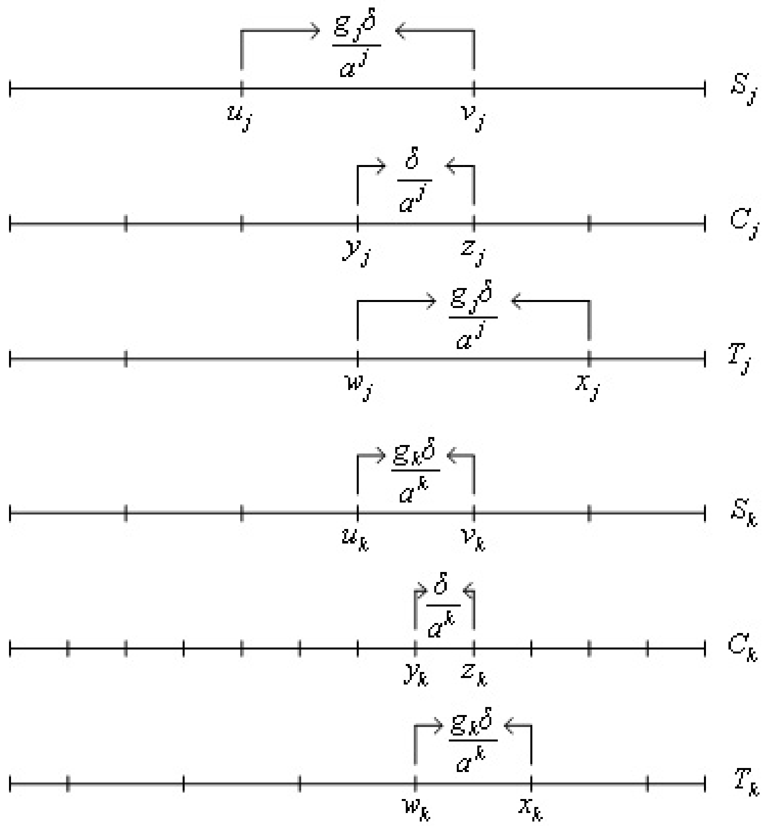


Figure 3.4: Central and side quantizers from MDSQ j and k.

descriptions for joint decoding. For example, the implicit central quantizer bin size for joint decoding of descriptions s_j and s_k are formulated as

$$\Delta_{s_j,s_k} = \min(v_j, v_k) - \max(u_j, u_k). \tag{3.17}$$

To compute the implicit central quantizer bin width for unbalanced joint decoding the relationship between the quantizer bin limits (minimum and maximum values) of the four side quantizers, S_j, S_k, T_j and T_k, needs to be formulated. With reference to Figure 3.4, for the MDSQ, $m = k$, the limits of the side quantizer bins for S_k, using Eq. (3.11) and Eq. (3.12) can be written as,

$$u_k = y_0 + (I_{k,0} + (1 - h_k)(1 - g_k))\frac{\delta}{a^k}, \tag{3.18}$$

$$v_k = u_k + g_k\frac{\delta}{a^k}. \tag{3.19}$$

Following Eq. (3.13) and Eq. (3.14), the limits of the side quantizer bins for T_k can be written as,

$$\begin{aligned} w_k &= y_0 + (I_{k,0} + h_k(1-g_k))\frac{\delta}{a^k}, && (3.20) \\ x_k &= w_k + g_k\frac{\delta}{a^k}. && (3.21) \end{aligned}$$

Similarly, the four limits for side quantizer bins for S_j and T_j are

$$\begin{aligned} u_j &= y_0 + (I_{j,0} + (1-h_j)(1-g_j))\frac{\delta}{a^j}, && (3.22) \\ v_j &= u_j + g_j\frac{\delta}{a^j}, && (3.23) \\ w_j &= y_0 + (I_{j,0} + h_j(1-g_j))\frac{\delta}{a^j}, && (3.24) \\ x_j &= w_j + g_j\frac{\delta}{a^j}. && (3.25) \end{aligned}$$

The relationship between the two index series, $I_{j,0}$ and $I_{k,0}$, for the two MDSQs is established as

$$I_{k,0} = a^{k-j}I_{j,0} + I_{k,j}, \tag{3.26}$$

where $I_{k,j} = \sum_{\alpha=j+1}^{k} a^{k-\alpha} i_\alpha$. Using Eq. (3.26), now the values of u_j, v_j, w_j and x_j can be rewritten in terms of a^k as

$$\begin{aligned} u_j &= y_0 + (I_{k,0} - I_{k,j} + a^{k-j}(1-h_j)(1-g_j))\frac{\delta}{a^k}, && (3.27) \\ v_j &= u_j + a^{k-j}g_j\frac{\delta}{a^k}, && (3.28) \\ w_j &= y_0 + (I_{k,0} - I_{k,j} + a^{k-j}h_j(1-g_j))\frac{\delta}{a^k}, && (3.29) \\ x_j &= w_j + a^{k-j}g_j\frac{\delta}{a^k}. && (3.30) \end{aligned}$$

The implicit central quantizer bin size for joint decoding of descriptions s_j and s_k is formulated as

$$\begin{aligned} \Delta_{s_j,s_k} &= \min(v_j, v_k) - \max(u_j, u_k), \\ &= \min\left((u_j + a^{k-j}g_j\frac{\delta}{a^k}), (u_k + g_k\frac{\delta}{a^k})\right) \\ &\quad - \max(u_j, u_k). && (3.31) \end{aligned}$$

By substituting Eq. (3.19) and Eq. (3.27) in Eq. (3.31), and choosing all different combinations of the values of $h_k = 0, 1$ and $h_j = 0, 1$ the minimum and maximum value of Δ_{s_j,s_k} is written as

$$\begin{aligned}\Delta_{s_j,s_{k_{\min}}} &= \frac{\delta}{a^k},\\ \Delta_{s_j,s_{k_{\max}}} &= g_k\frac{\delta}{a^k}.\end{aligned} \tag{3.32}$$

Similarly for other three combinations of joint decoding of unbalanced descriptions from different MDSQs, $m = j$ and $m = k$, the implicit central quantizer bin size can be formulated as

$$\Delta_{s_j,t_k} = \min(v_j, x_k) - \max(u_j, w_k), \tag{3.33}$$

$$\Delta_{t_j,s_k} = \min(x_j, v_k) - \max(w_j, u_k), \tag{3.34}$$

$$\Delta_{t_j,t_k} = \min(x_j, x_k) - \max(w_j, w_k). \tag{3.35}$$

One can show that the minimum and maximum values of the implicit central quantizer bin size for all combinations are,

$$\Delta_{s_j,s_{k_{\min}}} = \Delta_{s_j,t_{k_{\min}}} = \Delta_{t_j,s_{k_{\min}}} = \Delta_{t_j,t_{k_{\min}}} = \frac{\delta}{a^k}, \tag{3.36}$$

$$\Delta_{s_j,s_{k_{\max}}} = \Delta_{s_j,t_{k_{\max}}} = \Delta_{t_j,s_{k_{\max}}} = \Delta_{t_j,t_{k_{\max}}} = \frac{g_k\delta}{a^k}. \tag{3.37}$$

That means the smallest possible implicit central bin size for joint decoding of two unbalanced descriptions is the central bin size corresponding to the joint decoding of the two side descriptions from the higher index MDSQ (*i.e.*, $m = k$). The highest possible implicit central bin size for joint decoding of two unbalanced descriptions is the quantizer bin spread for the side quantizer of the higher index MDSQ, $m = k$.

3.2.2.3 Implicit Central Quantizer Bin Width Example

Figure 3.5 shows an example of the four side and their two implicit central quantizers using the side quantizer from same MDSQs and four implicit central quan-

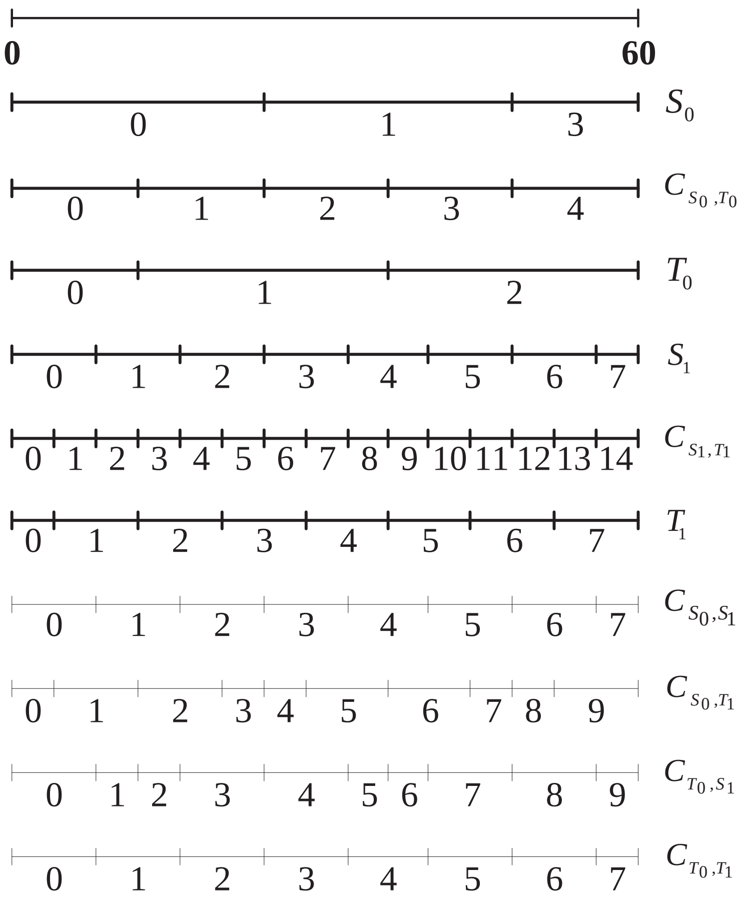


Figure 3.5: Implicit central and side quantizers from two MDSQs.

tizers from the side quantizer from different MDSQs. In Figure 3.5 S_0, T_0 and S_1, T_1 represents the side quantizers from the central quantizers C_{S_0,T_0} and C_{S_1,T_1} respectively. Similarly C_{S_0,S_1}, C_{S_0,T_1}, C_{T_0,S_0} and C_{T_0,T_1} represents the central quantizers from the side quantizer from different MDSQs. In Figure 3.5 the staggered case index assignment matrix with parameter $g = 2$ is used and the two central quantizers C_{S_0,T_0} and C_{S_1,T_1} are related to each other by factor $a = 3$. The minimum and maximum value of all the quantizers considered in Figure 3.5 is 0 and 60. Therefore, the central quantizer bin width for C_{S_0,T_0} and C_{S_1,T_1} are $\Delta_{S_0,T_0} = 12$ and $\Delta_{S_1,T_1} = 4$ respectively. As $g = 2$ is used the maximum side quantizer bin width for S_0, T_0 and S_1, T_1 are 24 and 8 respectively. By considering C_{S_0,T_1} the $\Delta_{S_0,T_1} = 4$ for $(0,0)$, $(0,3)$, $(1,3)$, $(1,6)$ and $(3,6)$ pairs of (S_0, T_1) and $\Delta_{S_0,T_1} = 8$ for $(0,1)$, $(0,2)$, $(1,4)$, $(1,5)$ and $(3,7)$ pairs of (S_0, T_1). It can easily be observed from Figure 3.5 that the implicit central quantizer bin width of central quantizers from side quantizers of different MDSQs *i.e.,* C_{S_0,S_1}, C_{S_0,T_1},

C_{T_0,S_0} and C_{T_0,T_1} is either 8 or 4. Those implicit central quantizer bin width values are equal to the bin width of the highest central or side quantizer, which is also formulated in Eq. (3.37) and Eq. (3.37) respectively.

3.2.2.4 Proposed Conditions

In the previous three subsections, the values of the implicit central quantizer from same and different MDSQs are formulated and simple example is discussed. In this section different conditions are proposed which should be satisfied to fulfil the distortion constraints explained in Section 3.2.1. For the conditions, any two MDSQs, $m = j$ and $m = k$, are considered where $k > j$, (an example is shown in Figure 3.4) from a set of MDSQs for deriving the design conditions for satisfying the distortion requirement (in Eq. (3.5)) for balanced joint decoding of two descriptions.

Proposition 1 *The sufficient and necessary condition for satisfying the joint distortion constraint for decoding two balanced descriptions from any two MDSQs, $m = j$ and $m = k$, where $k > j$, is $a > 1$, where a is the scaling factor for the bin spread for two consecutive side quantizers in the MDSQ.*

Proof : For satisfying the joint distortion constraint, in Eq. (3.5), the bin size of the implicit central quantizer for joint decoding for the MDSQ k must be smaller than the bin size of the implicit central quantizer for joint decoding for the MDSQ j, *i.e.*,

$$\begin{aligned} \Delta_{s_k,t_k} &< \Delta_{s_j,t_j}, \\ \frac{\delta}{a^k} &< \frac{\delta}{a^j}, \end{aligned} \tag{3.38}$$

which is simplified to $a^{k-j} > 1$. Since $k > j$, this is further simplified to $a > 1$. Therefore, the design condition is $a > 1$ for distortion constraint $D_{c_k} < D_{c_j}$.

■.

Now the individual decoding of description from two MDSQs is considered and the design conditions are formulated for meeting the distortion constraints in Eq. (3.1) - Eq. (3.2) as follows:

Proposition 2 *The sufficient and necessary condition for satisfying the side distortion constraints for two MDSQs, $m = j$ and $m = k$, where $k > j$, when individual descriptions are independently decoded is $a^{k-j}g_j > g_k$.*

Proof : For satisfying the side quantizer distortion constraints, Eq. (3.1) and Eq. (3.2), the bin spread of a side quantizer of the MDSQ, k, must be smaller than that of the corresponding side quantizer of the MDSQ, j, *i.e.,*

$$\frac{g_k\delta}{a^k} < \frac{g_j\delta}{a^j}, \tag{3.39}$$

which is simplified to $a^{k-j}g_j > g_k$. Therefore, the design condition is $a^{k-j}g_j > g_k$ for satisfying distortion constraints $D_{s_k} < D_{s_j}$ and $D_{t_k} < D_{t_j}$.

$\blacksquare$.

If one chooses $g_j = g_k = g_m$, then the condition $a^{k-j}g_j > g_k$ becomes the previously proposed condition $a^{k-j} > 1$.

Considering the above derivations, the design condition for meeting the distortion constraints in Eq. (3.7) for joint decoding of descriptions from different MDSQs are formulated as follows.

Proposition 3 *The sufficient condition for satisfying the joint distortion constraint for unbalanced descriptions from different MDSQs, $m = j$ and $m = k$, is $a^{k-j} \geq g_k$.*

Proof : For satisfying the joint distortion constraint, Eq. (3.7), the bin size of the implicit central quantizer for joint decoding of the two descriptions from different

MDSQs, $m = j$ and $m = k$, where $k > j$, must be smaller than the bin size of the implicit central quantizer for joint decoding of the two descriptions from the MDSQ $m = j$. That means for the four scenarios of choosing two unbalanced descriptions from the two MDSQs,

$$\begin{aligned} \Delta_{s_k,t_k} \leq \Delta_{s_j,s_{k_{\max}}}, \Delta_{s_j,t_{k_{\max}}}, \Delta_{t_j,s_{k_{\max}}}, \Delta_{t_j,t_{k_{\max}}} &\leq \Delta_{s_j,t_j}, \\ \frac{\delta}{a^k}\frac{g_k\delta}{a^k} &\leq \frac{\delta}{a^j}. \end{aligned} \tag{3.40}$$

This is simplified to $a^{k-j} \geq g_k$ and $g_k > 1$. Therefore, the design conditions are $a^{k-j} \geq g_k$ and $g_k > 1$ for satisfying distortion constraint $D_{c_j} \leq D_{c_{j,k}} \leq D_{c_j}$.

■.

Satisfying the above proposed three design conditions, $a > 1$, $a^{k-j}g_j > g_k$ and $a^{k-j} \geq g_k$, leads to

$$\frac{\delta}{a^k} \leq \Delta_{s_j,s_k}, \Delta_{s_j,t_k}, \Delta_{t_j,s_k}, \Delta_{t_j,t_k} \leq \frac{g_k\delta}{a^k} < \frac{\delta}{a^j} < \frac{g_j\delta}{a^j}. \tag{3.41}$$

The condition, $(\Delta_{s_j,s_k}, \Delta_{s_j,t_k}, \Delta_{t_j,s_k}, \Delta_{t_j,t_k}) \leq \frac{g_k\delta}{a^k}$ means that the bin size of the implicit central quantizer for joint decoding of the descriptions from different MDSQs, $m = j$ and $m = k$, is smaller than the bin size of an individual side description from the higher index MDSQ, $m = k$. This implies that the distortion due to joint decoding of the side quantizers from different MDSQs, $m = j$ and $m = k$, is less than the distortion due to individual decoding of a side description from the higher index MDSQ, $m = k$. Therefore, meeting the three design conditions jointly satisfies the joint unbalanced distortion constraint, Eq. (3.8).

These design conditions have been derived considering joint unbalanced decoding of descriptions from two different MDSQs. It can be extended to the joint unbalanced decoding of any number of descriptions coming from any number of MDSQs. For the two MDSQ case, the bin sizes of the implicit central quantizers for various combinations of joint decoding vary as shown in Eq. (3.41). For joint decoding, the implicit central quantizer bin size is a function of g_k, which is the index of the higher index MDSQ out of the two MDSQs, $m = j$ and $m = k$. One can easily show that for joint decoding of descriptions from more than two MDSQs, the implicit central quantizer bin size is a function of the index of the

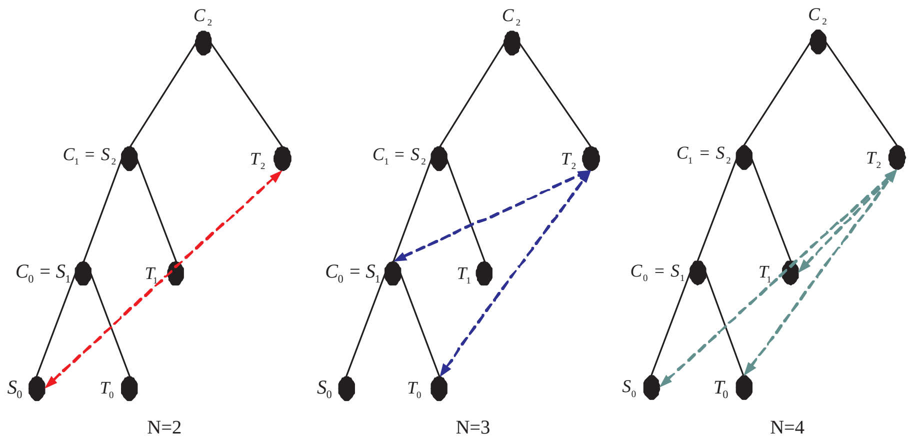


Figure 3.6: Hierarchical tree structure of successive side quantizer bin merging for three levels $J = 3$.

highest order MDSQ. For example, for the original J number of MDSQs with the central quantizers C_m, where $m = 0, \cdots, J-1$, the minimum and the maximum values of the bin size of the implicit central quantizer for joint decoding of descriptions are

$$\begin{aligned} \Delta_{s_0,t_0,\cdots,s_{J-1},t_{J-1}\min} &= \frac{\delta}{a^{J-1}}, \\ \Delta_{s_0,t_0,\cdots,s_{J-1},t_{J-1}\max} &= \frac{g_{J-1}\delta}{a^{J-1}}. \end{aligned} \tag{3.42}$$

3.3 Successive Side Quantizer Bin Merging :A Special Case Example

In the proposed MUDC framework, the central quantizer of the MDSQ m with a bin size of $\frac{\delta}{a^m}$ and the side quantizers with the bin size of $\frac{g_m\delta}{a^m}$ are considered, where the maximum side quantizer bin spread g_m is based on the index assignment matrix used and is calculated by Eq. (2.11) and Eq. (2.12) for staggered and modified nested index assignment matrix respectively. Now a special case is considered where $g_m = a$. Then the bin size for any of its side quantizers becomes $\frac{a\delta}{a^m}$, *i.e.*, $\frac{\delta}{a^{m-1}}$, which is the same as the bin size of the central quantizer of the MDSQ $m-1$. In this way one can consider a side quantizer of the higher index MDSQ as the central quantizer of the immediately preceding lower index MDSQ.

In other words, the set of MDSQs can be obtained by successive side quantizer bin merging (SSQBM). An example of hierarchy of MDSQs (for $J = 3$) is shown in Figure 3.6.

The hierarchical tree structure of the SSQBM shown in Figure 3.6 results in six side quantizers and descriptions. Let s_2 and t_2 be the descriptions created from the side quantizers, S_2 and T_2, respectively, derived from the central quantizer C_2, which is root of the tree. The number of quantizer bins of the central and side quantizers of the highest level of tree is selected according to the required joint rate-distortion constraints of that particular MDSQ. Let s_1 and t_1 be the two descriptions created from the side quantizers, S_1 and T_1, derived from the central quantizer C_1, which is the same as the side quantizer S_2. Similarly, s_0 and t_0 are the descriptions created from side quantizers, S_0 and T_0, respectively, derived from the central quantizer $C_0 = S_1$. The number of side quantizer bins depends on the index assignment matrix factor $g_m = a$ used in each tree level.

It is clear from Figure 3.6, that the same reconstructed value is obtained when descriptions from both children nodes and their parent node in the MDSQ tree hierarchy are jointly decoded. We can reconstruct the value with a smaller error, when two nodes from two different branches and levels of tree are received. Hence, it is more effective to transmit descriptions from the quantizers at different levels of the tree. Arrows with dotted lines in Figure 3.6 show possible combinations of the descriptions that are beneficial to combine in terms of improving the distortion performance when joint decoding of $N = 2, 3, 4$ descriptions, respectively. Another advantage of sending descriptions from different levels of the tree is obtaining progressive distortion improvement leading to quality scalability within the generalized MDC framework. That means the decoded value from a node of the lowest level of the tree is improved by joint decoding with any description from a higher level of the tree. The rate requirement of the description is increased if the selected descriptions are closer to the root of the tree because the number of quantizer indexes is also increased.

3.4 Simulation Results and Parameters

The results presented in this section is divided into three stages: Firstly considering the transmission under lossless channel conditions in order to study the rate-distortion performance and secondly considering the transmission under lossy channel conditions in order to study the robustness of the proposed MUDC using several MDSQs.

3.4.1 Performance under Lossless Channel Conditions

The performance of the proposed MUDC scheme under lossless channel conditions is divided into two parts. Firstly to verify the proposed conditions in Section 3.2.2 for different index assignment matrix parameters for each MDSQ and secondly the rate distortion performance of the MUDC for images.

3.4.1.1 Proposed Condition Verification for Distortion Constraints

In the first set of simulation results, the proposed design conditions in Section 3.2.2 are verified. Firstly, two MDSQs are considered to demonstrate the effect of different combinations of values for parameters g_j, g_k and $a > 1$, in terms of the individual and joint decoding quality. Table 3.1, Table 3.2 and Table 3.3 shows the individual decoding quality values ($D_{s_j}, D_{t_j}, D_{s_k}, D_{t_k}$) and Table 3.4, Table 3.5 and Table 3.6 shows the joint decoding quality values ($D_{c_{s_j,t_j}}$, $D_{c_{s_j,s_k}}$, $D_{c_{s_j,t_k}}$, $D_{c_{t_j,s_k}}$, $D_{c_{t_j,t_k}}$, $D_{c_{s_k,t_k}}$), in terms of the PSNR for the two MDSQs, $m = j$ and $m = k = j + 1$ for different parameter combinations. Results are shown for four test images (#1: *Barbara (on chair)*, #2: *Barbara (on floor)*, #3: *Gold Hill*, #4: *Black Board*).

Table 3.1 and Table 3.4 considers the same index assignment parameter $g_j = g_k = 2$ for the two MDSQs and $a = 2, 3$. Table 3.2 and Table 3.5 considers the same index assignment parameter $g_j = g_k = 3$ for the two MDSQs and $a = 3, 4$. Table 3.3 and Table 3.6 considers different index assignment parameters $g_j = 2$

Table 3.1: Individual decoding quality (PSNR in dB) and rate (in *bpp*) from two MDSQs having the same index assignment matrix parameters.

Test image #		Individual decoding MDSQ $m = j$				Individual decoding MDSQ $m = k$			
	a	R_{sj}	D_{sj}	R_{tj}	D_{tj}	R_{sk}	D_{sk}	R_{tk}	D_{tk}
1	2	0.7	29.9	0.7	30.0	1.1	31.9	1.1	34.2
1	3	0.7	29.9	0.7	30.0	1.5	34.2	1.5	36.6
2	2	0.7	29.5	0.7	29.5	1.1	31.2	1.1	34.0
2	3	0.7	29.5	0.7	29.5	1.4	34.4	1.4	36.2
3	2	0.3	30.0	0.3	30.0	0.7	32.9	0.7	33.3
3	3	0.3	30.0	0.3	30.0	0.9	34.2	0.9	34.3
4	2	0.3	33.9	0.3	34.2	0.5	35.5	0.5	37.5
4	3	0.3	33.9	0.3	34.2	0.7	38.0	0.7	38.3

and $g_k = 5$ for the two MDSQs and $a = 2, 5$. The increments of PSNR when the design constraints satisfy are shown in Table 3.7. The negative values correspond to the cases where the design conditions are not satisfied. For example, in the odd indexed data rows in Table 3.7 does not satisfy the constraints, $a^{k-j}g_j > g_k$ and $a^{k-j} \geq g_k$, therefore resulting in a decrease or an insignificant increase in the PSNR values, thus failing the distortion constraints. Results in these tables verify that satisfying the design conditions results in a significant increment in PSNR in the corresponding joint decoding for unbalanced descriptions.

Table 3.2: Individual decoding quality (PSNR in dB) and rate (in *bpp*) from two consecutive MDSQs having the same index assignment matrix parameters.

Test image #		Individual decoding MDSQ $m = j$				Individual decoding MDSQ $m = k$			
	a	R_{sj}	D_{sj}	R_{tj}	D_{tj}	R_{sk}	D_{sk}	R_{tk}	D_{tk}
1	3	0.7	27.7	0.7	27.8	1.5	32.7	1.5	34.9
1	4	0.7	27.7	0.7	27.8	1.7	33.5	1.7	36.6
2	3	0.7	27.3	0.7	27.4	1.4	32.3	1.4	35.2
2	4	0.7	27.3	0.7	27.4	1.7	32.7	1.7	37.2
3	3	0.3	26.2	0.3	31.2	0.9	30.2	0.9	35.4
3	4	0.3	26.2	0.3	31.2	1.2	32.8	1.2	36.4
4	3	0.3	31.9	0.3	32.4	0.7	36.0	0.7	36.8
4	3	0.3	31.9	0.3	32.4	0.9	37.0	0.9	39.0

Table 3.3: Individual decoding quality (PSNR in dB) and rate (in bpp) from two consecutive MDSQs having different index assignment matrix parameters.

Test image #				Individual decoding MDSQ $m = j$				Individual decoding MDSQ $m = k$			
	a	g_j	g_k	R_{sj}	D_{sj}	R_{tj}	D_{tj}	R_{sk}	D_{sk}	R_{tk}	D_{tk}
1	2	2	5	0.7	29.9	0.7	30.0	1.1	26.7	1.1	31.7
1	5	2	5	0.7	29.9	0.7	30.0	1.9	31.2	1.9	34.9
2	2	2	5	0.7	29.5	0.7	29.5	1.1	26.1	1.1	31.4
2	5	2	5	0.7	29.5	0.7	29.5	1.9	32.3	1.9	37.3
3	2	2	5	0.3	30.0	0.3	30.0	0.7	26.7	0.7	31.2
3	5	2	5	0.3	30.0	0.3	30.0	1.4	30.4	1.4	32.3
4	2	2	5	0.3	33.9	0.3	34.2	0.5	30.5	0.5	35.6
4	5	2	5	0.3	33.9	0.3	34.2	1.1	35.2	1.1	39.1

Table 3.4: Joint decoding quality (PSNR in dB) and rate (in bpp) from two MDSQs having the same index assignment matrix parameters.

Test image #		Joint decoding balanced				Joint decoding unbalanced							
	a	$R_{c_{sj,tj}}$	$D_{c_{sj,tj}}$	$R_{c_{sk,tk}}$	$D_{c_{sk,tk}}$	$R_{c_{sj,sk}}$	$D_{c_{sj,sk}}$	$R_{c_{sj,tk}}$	$D_{c_{sj,tk}}$	$R_{c_{tj,sk}}$	$D_{c_{tj,sk}}$	$R_{c_{tj,tk}}$	$D_{c_{tj,tk}}$
1	2	1.4	32.3	2.2	35.6	1.8	34.8	1.8	35.6	1.8	34.9	1.8	35.6
1	3	1.4	32.3	3.0	37.8	2.2	37.3	2.2	36.9	2.2	37.3	2.2	36.9
2	2	1.4	31.8	2.2	35.5	1.8	34.4	1.8	35.5	1.8	34.5	1.8	35.5
2	3	1.4	31.8	2.8	37.8	2.1	36.9	2.1	36.6	2.1	37.0	2.1	36.6
3	2	0.6	31.7	1.4	35.1	1.0	34.6	1.0	35.0	1.0	34.7	1.0	35.1
3	3	0.6	31.7	1.8	36.2	1.5	36.0	1.5	36.0	1.5	36.1	1.5	36.1
4	2	0.6	36.0	1.0	38.2	0.8	38.0	0.8	38.3	0.8	37.8	0.8	38.3
4	3	0.6	36.0	1.4	40.0	1.0	39.4	1.0	39.3	1.0	39.4	1.0	39.3

Table 3.5: Joint decoding quality (PSNR in dB) and rate (in bpp) from two consecutive MDSQs having the same index assignment matrix parameters.

Test image #		Joint decoding balanced				Joint decoding unbalanced							
	a	$R_{c_{sj,tj}}$	$D_{c_{sj,tj}}$	$R_{c_{sk,tk}}$	$D_{c_{sk,tk}}$	$R_{c_{sj,sk}}$	$D_{c_{sj,sk}}$	$R_{c_{sj,tk}}$	$D_{c_{sj,tk}}$	$R_{c_{tj,sk}}$	$D_{c_{tj,sk}}$	$R_{c_{tj,tk}}$	$D_{c_{tj,tk}}$
1	3	1.4	32.3	3.0	37.8	2.2	36.3	2.2	36.7	2.2	35.9	2.2	37.0
1	4	1.4	32.3	3.4	39.0	2.4	37.5	2.4	37.1	2.4	37.4	2.4	37.1
2	3	1.4	31.8	2.8	37.8	2.1	35.5	2.1	36.6	2.1	35.4	2.1	36.8
2	4	1.4	31.8	3.4	39.4	2.4	37.3	2.4	37.3	2.4	37.3	2.4	37.3
3	3	0.6	31.7	1.8	36.2	1.2	35.5	1.2	35.8	1.2	35.5	1.2	36.3
3	4	0.6	31.7	2.4	38.0	1.8	36.3	1.8	37.0	1.8	36.4	1.8	37.1
4	3	0.6	36.0	1.4	40.0	1.0	39.3	1.0	38.9	1.0	38.9	1.0	39.2
4	3	0.6	36.0	1.8	40.6	1.2	39.5	1.2	39.7	1.2	39.6	1.2	39.6

Table 3.6: Joint decoding quality (PSNR in dB) and rate (in bpp) from two consecutive MDSQs having different index assignment matrix parameters.

Test image #				Joint decoding balanced				Joint decoding unbalanced							
	a	g_j	g_k	$R_{c_{sj,tj}}$	$D_{c_{sj,tj}}$	$R_{c_{sk,tk}}$	$D_{c_{sk,tk}}$	$R_{c_{sj,sk}}$	$D_{c_{sj,sk}}$	$R_{c_{sj,tk}}$	$D_{c_{sj,tk}}$	$R_{c_{tj,sk}}$	$D_{c_{tj,sk}}$	$R_{c_{tj,tk}}$	$D_{c_{tj,tk}}$
1	2	2	5	1.4	32.3	2.2	35.3	1.8	31.5	1.8	34.6	1.8	34.3	1.8	34.9
1	5	2	5	1.4	32.3	3.8	39.8	2.6	36.6	2.6	37.4	2.6	36.1	2.6	38.1
2	2	2	5	1.4	31.8	2.0	35.0	1.7	31.2	1.7	34.4	1.7	33.9	1.7	34.8
2	5	2	5	1.4	31.8	3.8	39.8	2.6	35.7	2.6	38.0	2.6	35.7	2.6	38.0
3	2	2	5	0.6	31.7	1.4	35.0	1.0	31.9	1.0	34.3	1.0	34.2	1.0	34.5
3	5	2	5	0.6	31.7	2.8	38.6	1.7	36.2	1.7	36.9	1.7	36.0	1.7	37.1
4	2	2	5	0.6	36.0	1.0	38.2	0.8	35.2	0.8	38.0	0.8	37.2	0.8	38.0
4	5	2	5	0.6	36.0	2.2	40.4	1.7	38.2	1.7	40.4	1.7	38.3	1.7	40.2

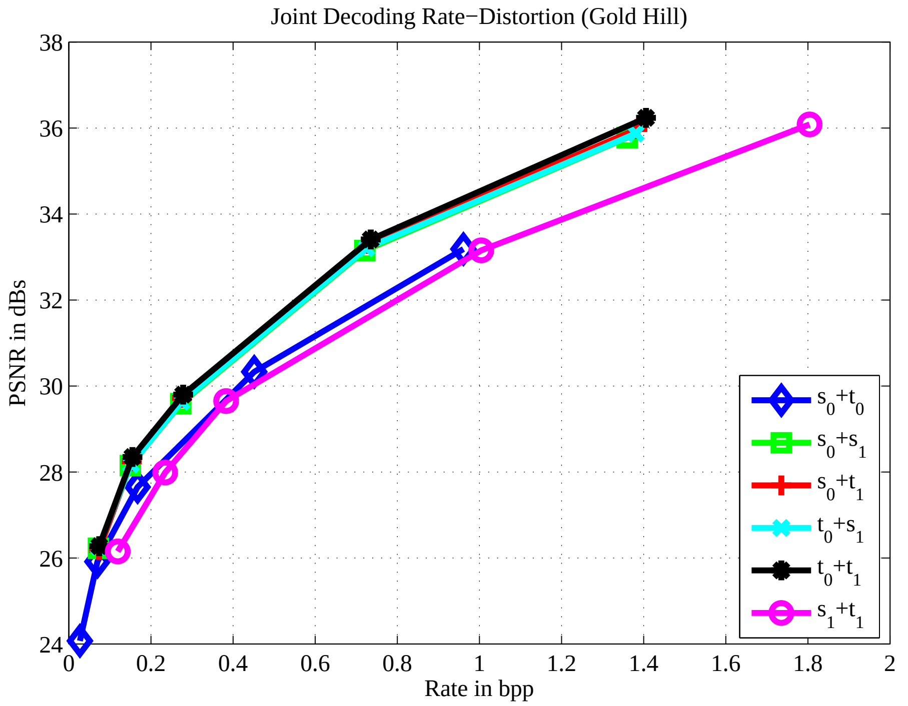
Table 3.7: Difference of individual and joint decoding quality (PSNR in dB) from two consecutive MDSQs having different index assignment matrix parameters.

Test image #				Individual PSNR Difference		Unbalanaced joint PSNR Difference w.r.t $D_{c_{s_j},t_j}$			
	a	g_j	g_k	$D_{s_k} - D_{s_j}$	$D_{t_k} - D_{t_j}$	$D_{c_{s_j},s_k}$	$D_{c_{s_j},t_k}$	$D_{c_{t_j},s_k}$	$D_{c_{t_j},t_k}$
1	2	2	5	−3.2	1.7	−0.8	2.3	2.2	2.6
1	5	2	5	1.3	4.9	4.3	5.1	3.8	5.8
2	2	2	5	−3.4	0.9	−0.6	2.6	2.1	3.0
2	5	2	5	2.8	7.8	3.9	6.2	3.9	6.2
3	2	2	5	−3.3	1.2	−0.2	2.6	2.5	2.8
3	5	2	5	0.4	2.3	4.5	5.2	4.3	5.4
4	2	2	5	−3.4	1.4	−0.8	2.0	1.2	2.0
4	5	2	5	1.3	4.9	2.2	4.4	2.3	4.2

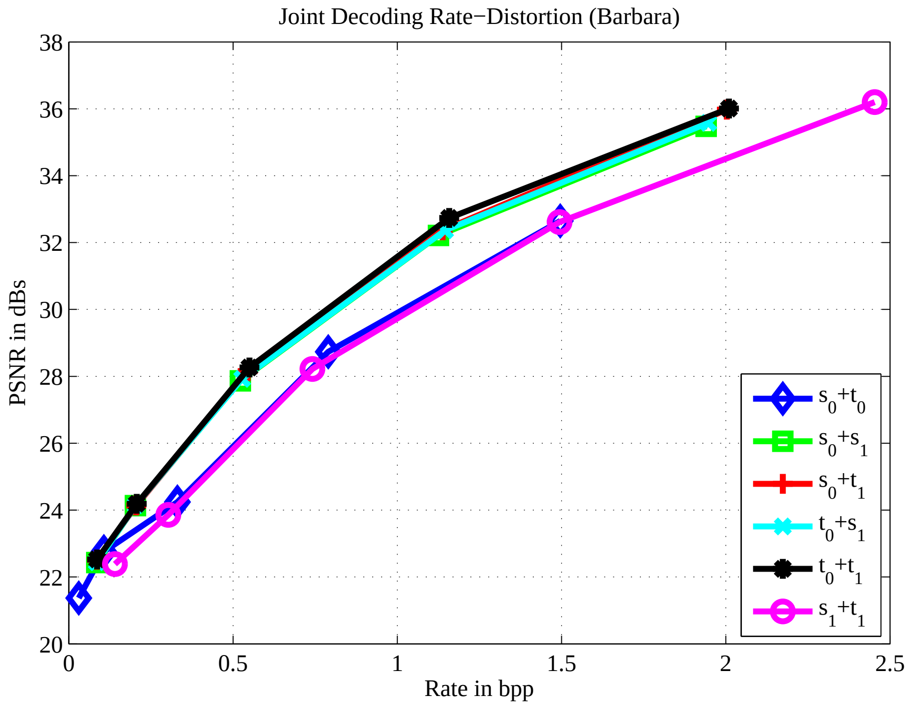
3.4.1.2 Rate Distortion Performance

In the second set of simulation results, the rate-distortion performance of MUDC scheme using $J = 2$ MDSQs, that generates four descriptions *i.e.,* (s_0, t_0, s_1, t_1) is presented. The quantizer bin size of the central quantizer for the two MDSQs are related to each other by the factor $a = 2$. If the central quantizer bin size of first MDSQ is δ then the central quantizer bin size of the second MDSQ is $\frac{\delta}{a}$. For the rate-distortion graphs the same index assignment matrix parameters is used for both MDSQs *i.e.,* ($g_0 = g_1 = 2$). The rate-distortion curves for joint decoding of N descriptions for *Gold Hill*, *Barbara (on chair)*, *Barbara (on floor)* and *Blackboard* images are shown in Figure 3.7 and Figure B.1 for $N = 2$. Similarly the rate-distortion curves for joint decoding of N descriptions for *Gold Hill*, *Barbara (on chair)*, *Barbara (on floor)* and *Blackboard* images are shown in Figure 3.8 and Figure B.2 for $N \in \{3, 4\}$. From Figure 3.7 and Figure B.1, it is evident that for $N = 2$, the joint decoding rate-distortion performance of the descriptions from different MDSQs is better than that of the descriptions chosen from the same MDSQ. For $N = 3$, the joint decoding rate-distortion performance is not the same for all combinations. For some side quantizers there is no refinement because the bins totally fall within the bins of the other side quantizers. Therefore, few combinations for $N = 3$ give better rate-distortion performance than that for the other combinations as also observed from Figure 3.8 and Figure B.2. The joint decoding distortion for $N = 4$ is the same as the joint decoding distortion of the descriptions from MDSQ having smaller central quantizer bin size. There is an increase in the rate (adding more redundancy) as the number of descriptions is increased to four, thereby, resulting in overall high robustness of the system as explained in the following subsection where the performance of the proposed scheme is evaluated under noisy channel conditions.

In the final set of simulations under lossless conditions, the rate-distortion performance of different combinations of the descriptions is compared using a similar sort of tree structure used in the SSQBM scheme explained in Section 3.3. Figure 3.9, Figure 3.10 and Figure 3.11 shows the joint decoding rate-distortion performance of *Gold Hill* and *Barbara (on chair)* image for $N = 2$, $N = 3$ and $N = 4$ descriptions from $J = 3$ MDSQs respectively. For $N = 2$ descriptions it is more beneficial in terms of the rate-distortion performance to select one description from the highest and another from the lowest level of the tree. Similarly

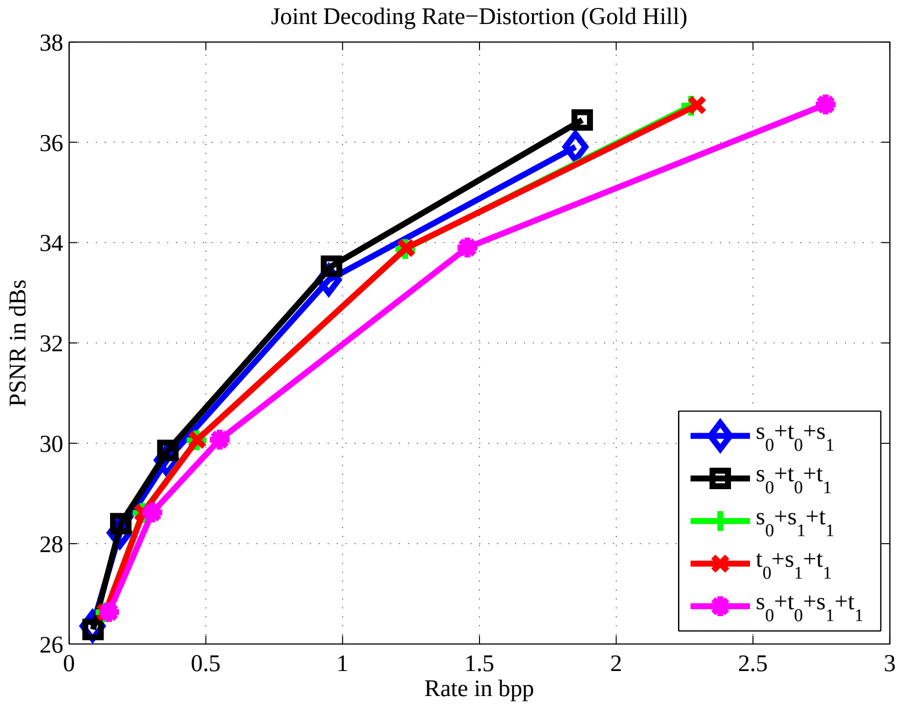


(a)

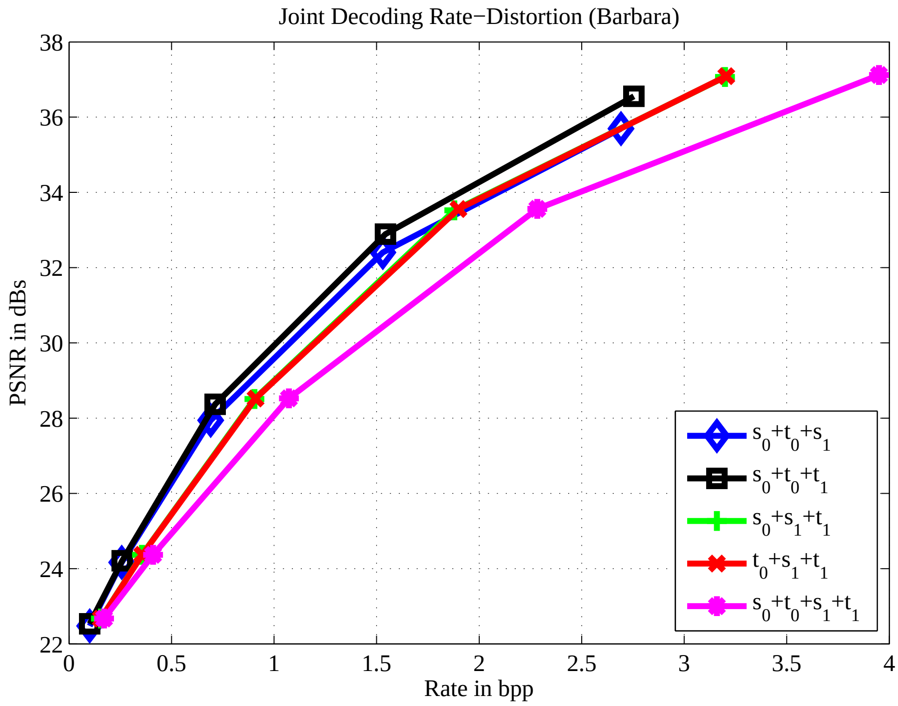


(b)

Figure 3.7: Joint decoding rate-distortion curves for $N = 2$ descriptions from $J = 2$ MDSQs for (a) *Gold Hill* (b) *Barbara (on chair)* images.

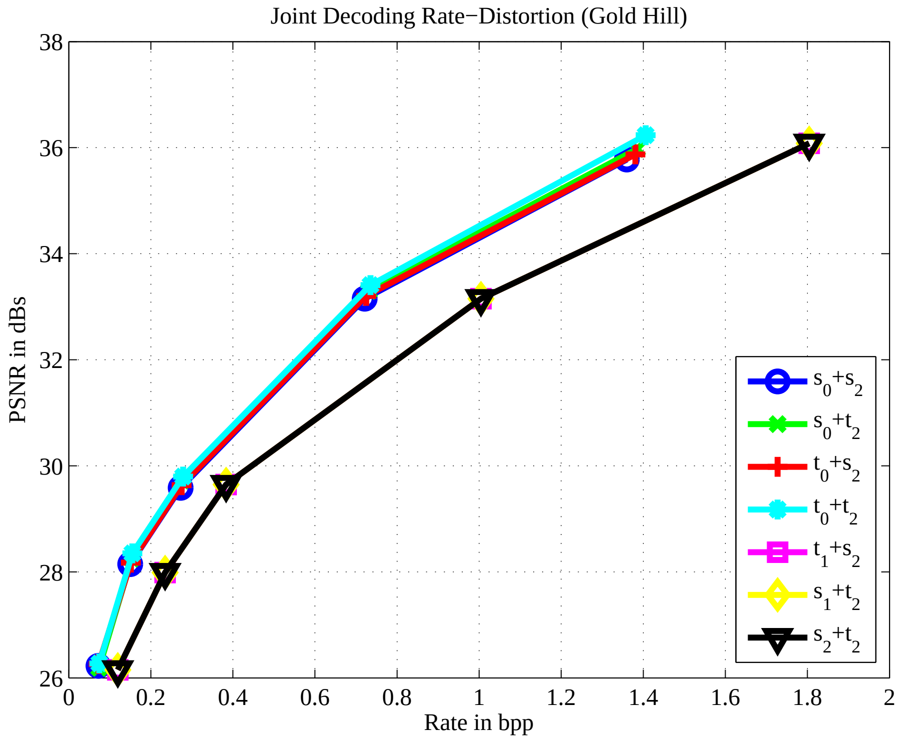


(a)

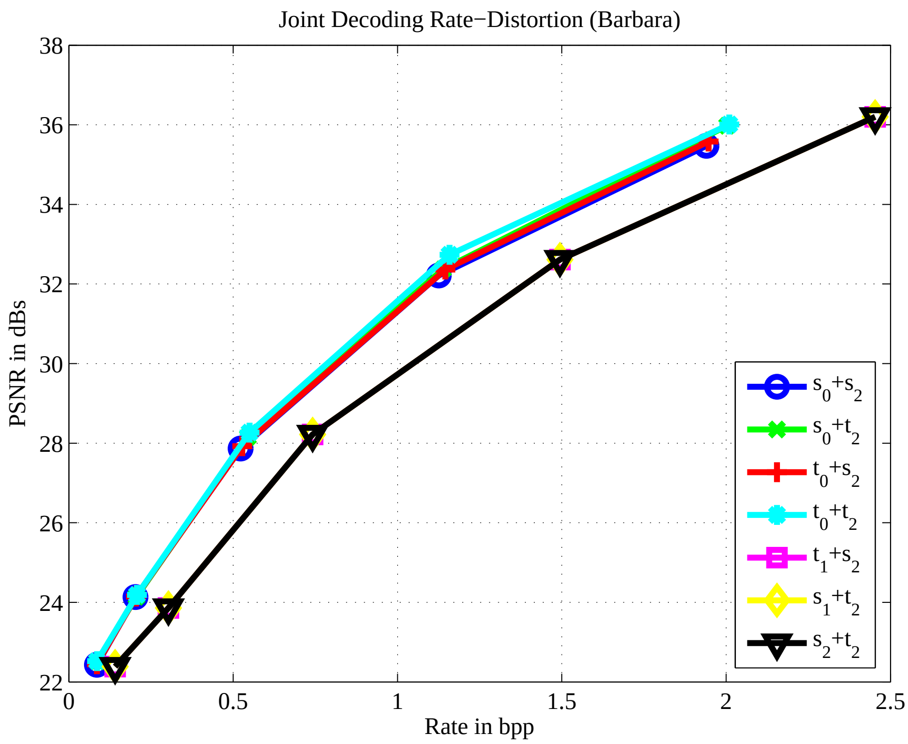


(b)

Figure 3.8: Joint decoding rate-distortion curves for $N = 3, 4$ descriptions from $J = 2$ MDSQs for (a) *Gold Hill* (b) *Barbara (on chair)*.

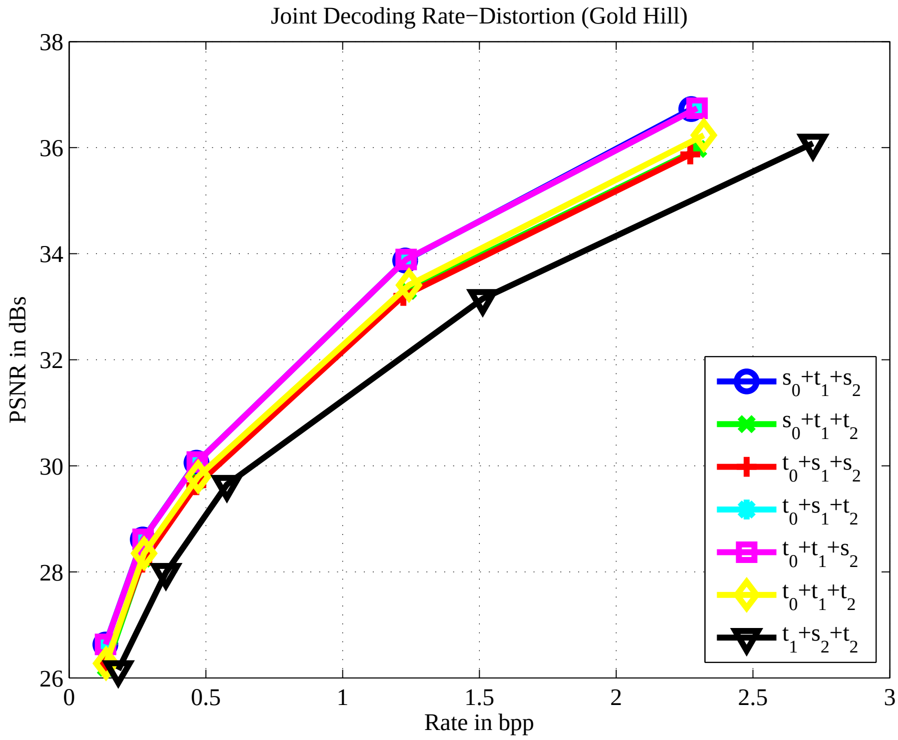


(a)

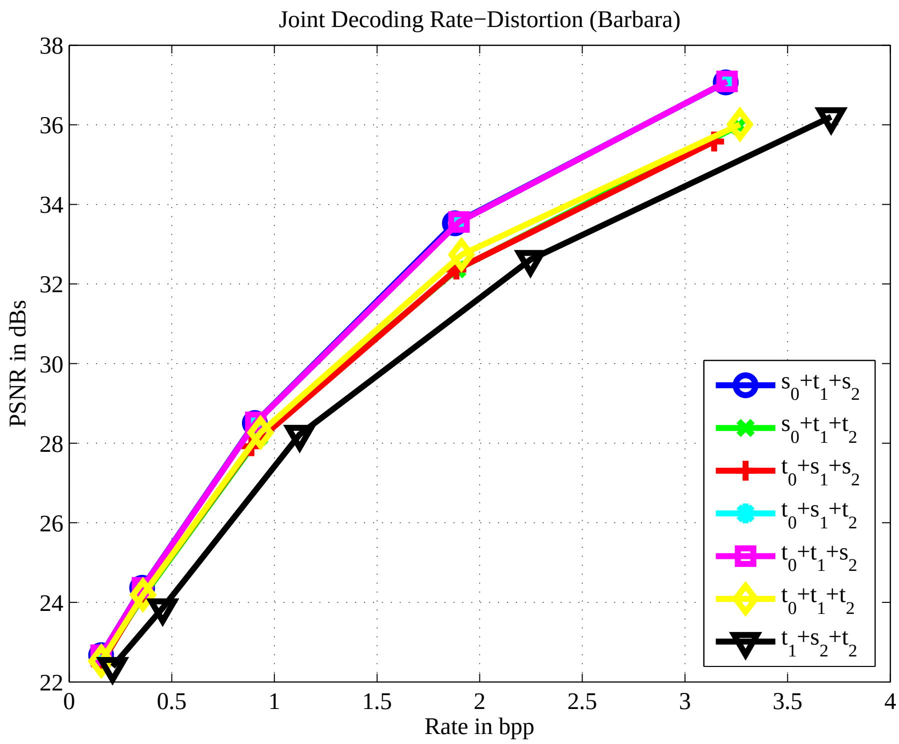


(b)

Figure 3.9: Joint decoding rate-distortion curves for $N = 2$ descriptions from $J = 3$ MDSQs for (a) *Gold Hill* (b) *Barbara (on chair)* images.

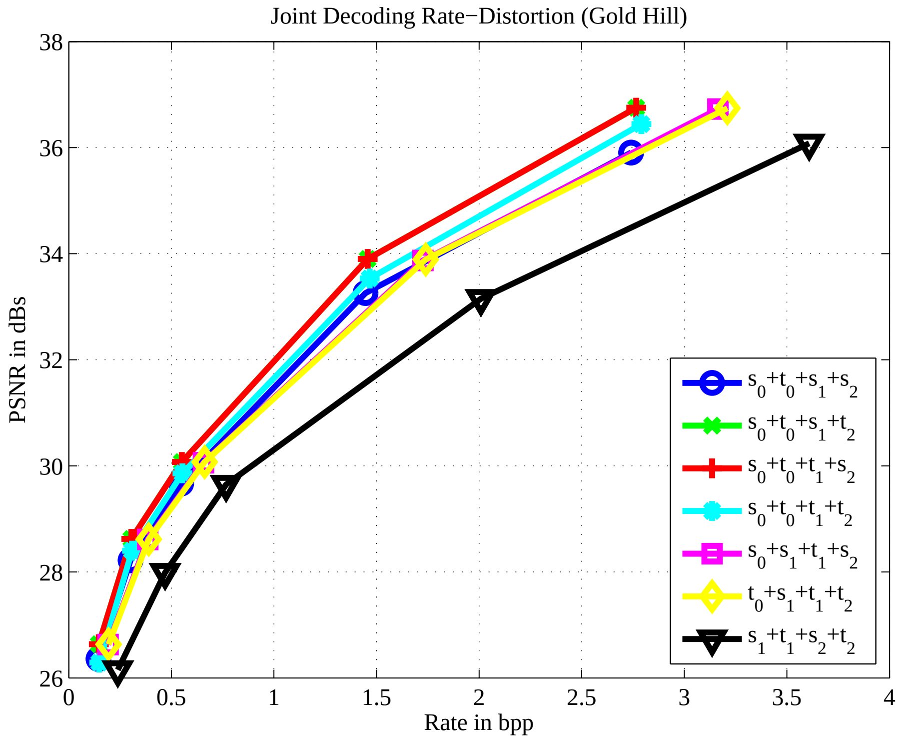


(a)

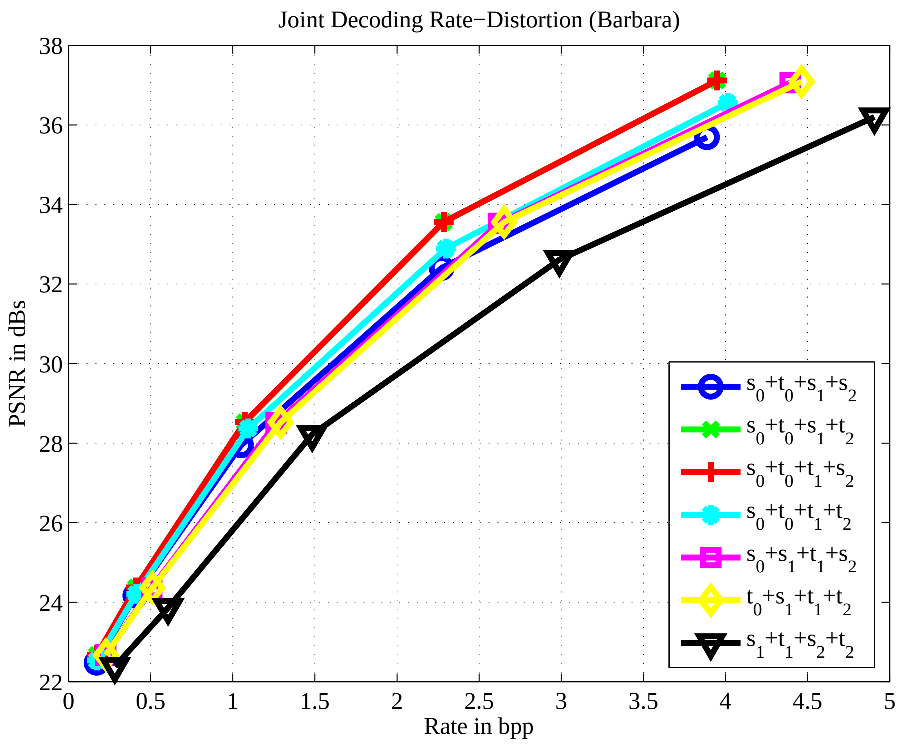


(b)

Figure 3.10: Joint decoding rate-distortion curves for $N = 3$ descriptions from $J = 3$ MDSQs for (a) *Gold Hill* (b) *Barbara (on chair)* images.



(a)



(b)

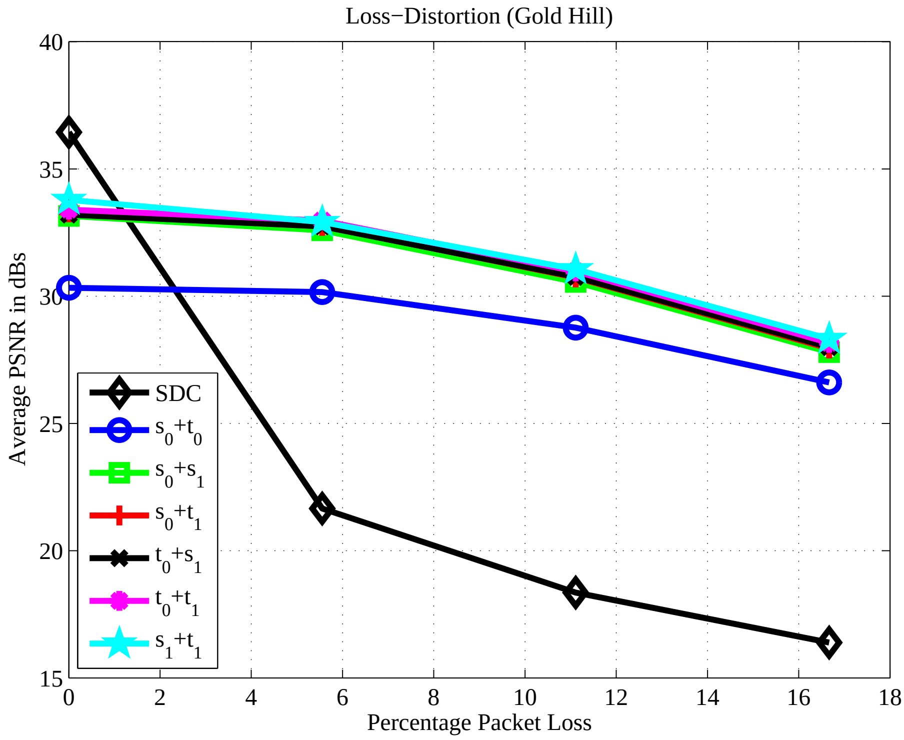
Figure 3.11: Joint decoding rate-distortion curves for $N = 4$ descriptions from $J = 3$ MDSQs (a) *Gold Hill* (b) *Barbara (on chair)* images.

for $N = 3$ it is more beneficial in terms of the rate-distortion performance to select descriptions from different levels and branches of the tree. For $N = 4$ descriptions it is better to send more descriptions from lower levels of the tree than sending descriptions from upper levels. These rate distortion plots in Figure 3.9, Figure 3.10 and Figure 3.11 show that the quality scalability can be achieved if the descriptions from different levels of the tree are transmitted through different paths. Generating more than two descriptions using such a kind of tree structure not only provides the quality scalability in terms of increase in descriptions, but also increase the overall robustness of the system as the number of descriptions is increased, which is demonstrated in the following section.

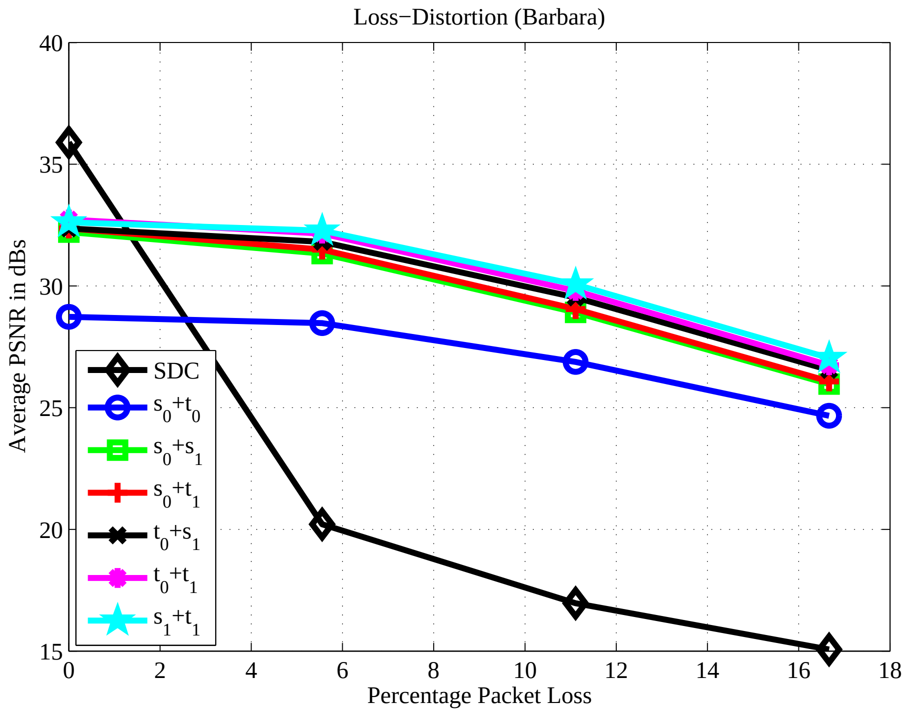
3.4.2 Performance under Lossy Channel Conditions

This set of simulations evaluates the performance of the proposed MUDC scheme under lossy channel conditions. In this case, the transmission of each description over a packet erasure channel is evaluated. Each side description data is packetized and different packet loss percentages are considered. Let M be the total number of packets for each description and l be the number of lost packets resulting in a total of ${}^{M}C_{l}$ number of combinations to loose l packets from the total of M packets. The average PSNR at a particular number of packet loss rate is then calculated by measuring and averaging the PSNR values of all possible packet loss combinations at that packet loss rate. For this set of experiments the packet erasure channel model having a loss rate varying between 0% and 18% is considered.

The robustness performance, in terms of the average PSNR, of the joint decoded image for multi-channel unbalanced MDSQs and individual decoding of single description coding (SDC) at different packet loss rates is shown in Figure 3.12 and Figure 3.13 for *Gold Hill* and *Barbara (on chair)* images. For MUDC, $J = 2$ MDSQs are considered and joint decoding for $N = 2$, $N = 3$ and $N = 4$ descriptions is evaluated under packet erasure channel. The SDC is the same as conventional wavelet based progressive coding. In the case of no packet loss, the SDC provides better performance compared to all different combinations of joint decoding in the proposed MUDC scheme. This is mainly due to the rate

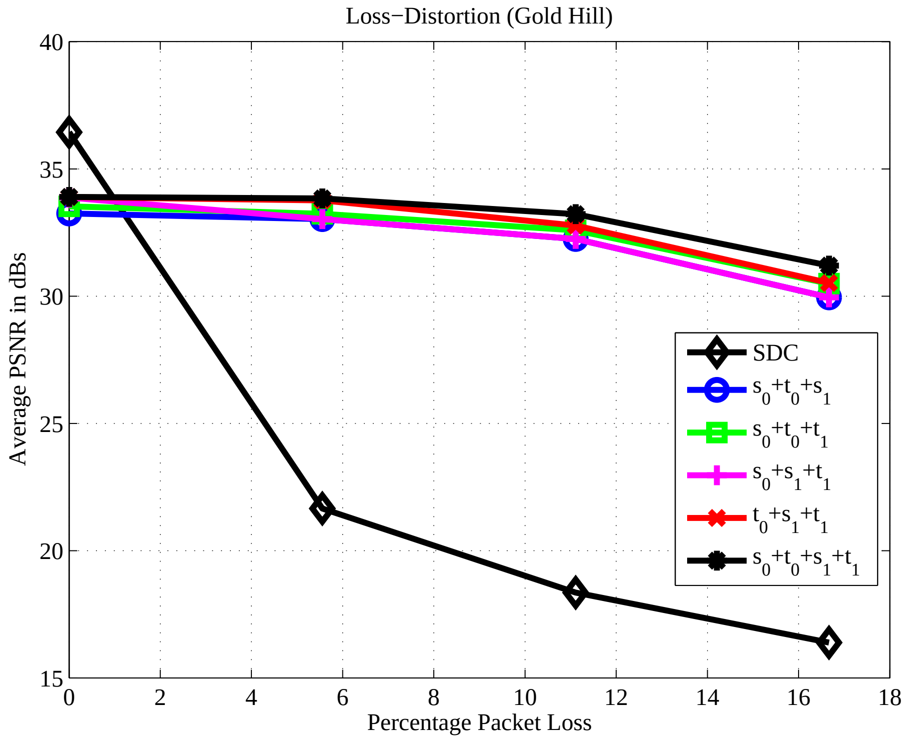


(a)

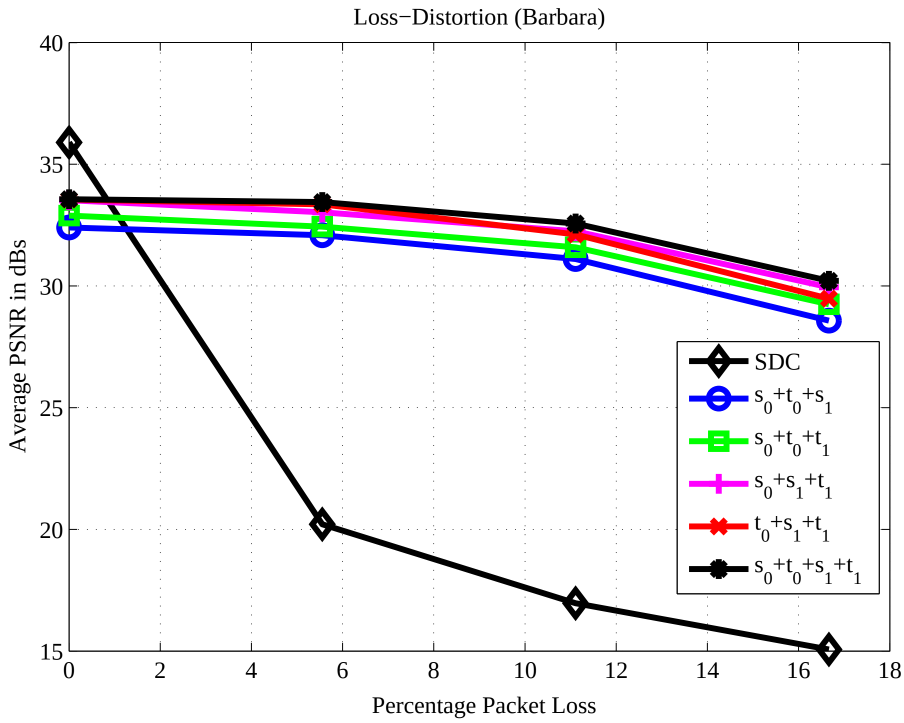


(b)

Figure 3.12: Effect of packet loss on PSNR for different percentage of packet drops for $N = 2$ descriptions from $J = 2$ MDSQs for (a) *Gold Hill* (b) *Barbara (on chair)* images.



(a)



(b)

Figure 3.13: Effect of packet loss on PSNR for different percentage of packet drops for $N = 3, 4$ descriptions from $J = 2$ MDSQs for (a) *Gold Hill* (b) *Barbara (on chair)* images.

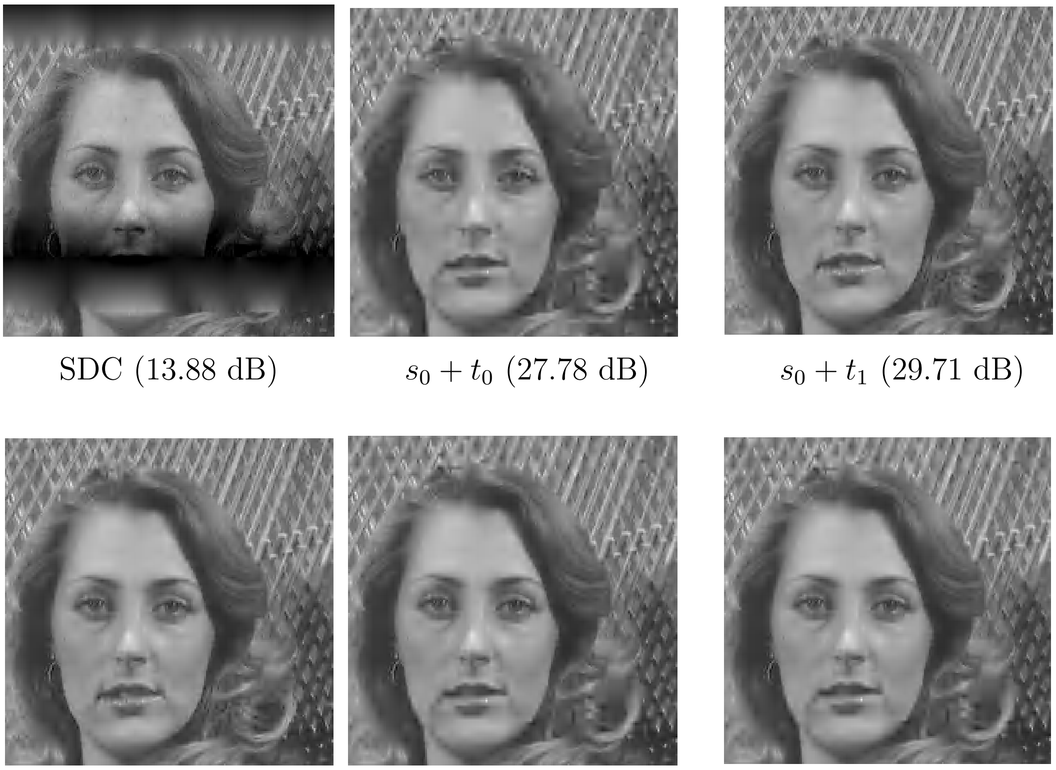
SDC (17.12 dB) $s_0 + t_0$ (29.72 dB) $s_0 + t_1$ (31.49 dB)

$t_0 + s_1$ (31.89 dB) $s_0 + t_0 + t_1$ (32.59 dB) $s_0 + t_0 + s_1 + t_1$ (33.05 dB)

Figure 3.14: A part of the decoded *Gold Hill* image after 18% packet loss in each description for different joint description decoding scenarios.

increment due to multiple descriptions used in the MDC-based solutions, which can be regarded as the cost of making the descriptions robust. However, the performance of SDC scheme deteriorates rapidly even for smaller percentages of packet losses. The loss-distortion plots for the joint decoding for $N = 2$ from different MDSQ is almost the same as the loss-distortion curves of the joint decoding of the MDSQ with smaller central quantizer bin size. Figure 3.13 shows the loss-distortion curves for joint decoding of $N = 3$ and $N = 4$ side descriptions. The loss-distortion plots become more horizontal, *i.e.,* more robust, as the number of descriptions jointly decoded is increased.

Figure 3.14 shows a portion of the decoded *Gold Hill* image and their PSNR value in dB for SDC and the MUDC scheme for $J = 2$ and $N = 2, 3, 4$ description joint decoding for various joint decoding scenarios with 18% packet loss in each description. The superior visual quality and increase in the PSNR value of the decoded image is easily observed as the number of descriptions is increased. Similar visual quality and PSNR values improvement can be observed for the



SDC (13.88 dB) $s_0 + t_0$ (27.78 dB) $s_0 + t_1$ (29.71 dB)

$t_0 + s_1$ (30.92 dB) $s_0 + t_0 + t_1$ (31.08 dB) $s_0 + t_0 + s_1 + t_1$ (31.82 dB)

Figure 3.15: A part of the decoded *Barbara (on chair)* image after 22% packet loss in each description for different joint description decoding scenarios.

Barbara (on chair) image in Figure 3.15 when 22% of packets are dropped from each description. Both in Figure 3.14 and Figure 3.15, equal percentage of packet loss is considered in each description. In Figure 3.16 comparison of the decoded image and their PSNR values is presented when different percentage of packet drops are considered in each description for *Barbara (on floor)* image. Figure 3.16 (a) shows the decoded image when 22% packets are lost in description s_0 and s_1 and 11% packets are lost in descriptions t_0 and t_1. On the otherhand, Figure 3.16 (b) considers 22% packets are dropped from each description.

The PSNR values of the decoded image from different percentage of packet loss in each description is 0.2 dB better for $N = 2$ and 0.5 dB for $N = 3, 4$ when compared with the decoded image having equal percentage of packet drops in each description.

$s_0 + t_0$ (28.06 dB)

$s_0 + t_0$ (27.87 dB)

$s_0 + t_1$ (29.97 dB)

$s_0 + t_1$ (29.58 dB)

$s_0 + t_0 + t_1$ (31.50 dB)

$s_0 + t_0 + t_1$ (31.08 dB)

$s_0 + t_0 + s_1 + t_1$ (32.07 dB)

$s_0 + t_0 + s_1 + t_1$ (31.69 dB)

Figure 3.16: A part of the decoded *Barbara (on floor)* image for different joint description decoding scenarios after 22% from s_0, s_1 and 11% packet loss from t_0, t_1.

3.5 Summary

In this chapter, a novel scheme for generalized MUDC resulting in joint decoding of descriptions with balanced and unbalanced rate-distortion performances was presented. The hierarchically defined MDSQs are used in the proposed framework by the successive refinement of the central quantizers. The main parameters involved in MUDC were the central quantizer refinement factor, a and the index assignment matrix parameter g_m for the m^{th} MDSQ in the hierarchy. The sufficient and necessary conditions for meeting the required rate-distortion constraints were proposed for single and joint decoding of descriptions from two MDSQs, $m = j$ and $m = k$, where $k > j$, as $a > 1$, $a^{k-j}g_j > g_k$ and $a^{k-j} \geq g_k$. The proposed conditions are also verified by demonstrating the PSNR increment for different combinations of parameter values. It is observed that at the same rate, the unbalanced joint decoding gives 1.1 dB better performance than the balanced joint decoding.

An efficient realization of the scheme was also shown by using the successive side quantizer bin merging of the initial MDSQ. The rate-distortion performance for joint decoding of $N = 2$ descriptions has shown that the unbalanced descriptions coming from different MDSQs resulted in superior performance compared to the balance descriptions coming from the same MDSQ. It is also shown that the quality scalability can be achieved if the joint decoding descriptions are selected from different levels and branches of the SSQBM tree. The flexibility to add and remove redundancy in terms of number of descriptions is also achieved by using SSQBM tree structure.

Chapter 4

Multiple Description Scalar Quantizer with Successive Refinement

This chapter focuses on achieving quality scalability in MDC framework by successive refinement of the individual description. The proposed method for achieving quality scalability starts with MDSQ-based MDC for the base layer and then successively refines the side quantizer to design a new framework called MDSQ-SR. The objective of the MDSQ-SR design is to improve the distortion for every refinement layer of a side description when individually decoded and for any combination of levels of refinement of the two refined side descriptions for joint decoding. The rest of the chapter is organized as follows: An overview of the scalable MDC scheme is presented in Section 4.1. The MDSQ-SR design, its distortion constraints and proposed conditions for successful decoding are presented in Section 4.2. Simulation results using the proposed scheme under both balanced and unbalanced description scenarios with application to quality scalable image coding are shown in Section 4.3 followed by the summary in Section 4.4.

4.1 Background

The emergence of using the wavelet transform in image coding has resulted in incorporating extra features, such as scalable decoding into image coding algorithms. As scalable coding usually uses hierarchical representations of spatial-quality coding layers with progressive interdependencies, any error in lower layers, for example in low frequency subbands, can propagate into the higher layers. Therefore, in scalable coding, low spatial-quality layers need to be highly protected for channel errors. In addition to hierarchical channel coding strategies, MDC can also be used to make scalable coded bit stream robust. One such example includes EMDSQ, where a set of side quantizers generating two descriptions are derived from an embedded central quantizer [53, 54]. Other examples either avoid using MDSQ [106] or use MDSQ only as the base layer in an MDC system with a single enhancement layer [107].

Early MDSQ algorithms focussed on obtaining descriptions with balanced rate-distortion performance. In Chapter 3 the conditions for obtaining unbalanced descriptions and their joint decoding are derived and also these conditions are extended for creating more than two descriptions for MDC. In this chapter, inspired from the results of MUDC scheme, a framework for successive refinement of side quantizers of the MDSQ is formulated to obtain progressive quality update for side quantizers.

In contrast to EMDSQ, the MDSQ-SR design considers different index assignment matrices (resulting into non-overlapped and overlapped side quantizer bins) to incorporate different amounts of redundancy between the descriptions at the base layer. The side quantizer bins of the base layer are then successively refined to guarantee the individual and joint distortion reduction for the enhancement layers. Using different index assignment matrices at the base layer facilitates the user to incorporate different amounts of redundancy among the descriptions depending on the number of diagonals filled in the index assignment matrix. The amount of redundancy among the descriptions for each enhancement layer is controlled by the refinement factor of the side quantizer bins. In MDSQ-SR design, different strategies for quantizer bin index assignments, such as staggered index assignment resulting in non-overlapped side quantizer bins and modified

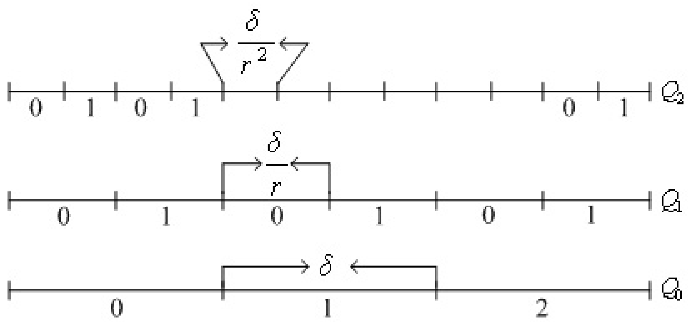


Figure 4.1: Embedded quantizer for three levels $P = 3$.

nested index assignment resulting in overlapped quantizer bins are considered. This chapter presents the conditions that the side quantizer bin sizes and the refinement factors have to meet in order to ensure progressive quality increments for a given side quantizer set, as well as for the central quantizer that corresponds to joint decoding scenarios.

4.2 MDSQ-SR Design

In embedded quantization, the quantizer bins at higher data rates are embedded within the quantizer bins of lower data rates. In other words, each quantizer bin of lower rates is refined or split into smaller quantizer bins for higher data rates to refine the previous quantized value.

In MDSQ-SR the main aim is to create two scalable side descriptions which are capable of joint decoding using equal and unequal number of refinement levels from each of the side description leading to balanced and unbalanced multiple description scenario respectively. For this case, we start with a central quantizer and two corresponding side quantizers leading to a base layer and then successively refine each of the description from the side quantizers to guarantee successive distortion reductions leading to enhancement layers for side quantizers. For the base layer, different index assignment strategies are considered depending on redundancy between the descriptions, such as staggered index assignment resulting in non-overlapped side quantizer bins and modified nested index assignment resulting in overlapped quantizer bins. The advantage of using different index assignment matrices facilitates the user to add required amounts of redundancy at the base layer. The redundancy between the descriptions for the enhancement layer is controlled by refinement factor of the side quantizers. The design conditions of the MDSQ-SR are derived for joint and individual decoding distortion constraints of the side descriptions for different refinement layers.

4.2.1 Design Distortion Constraints

For an embedded quantizer, let P be the number of levels of embedded layers, $Q_0, Q_1, \cdots, Q_{P-1}$. Then the quantizer bins of Q_p are embedded within the quantizer bins of Q_{p-1}, where p is the embedded level index. In other words, each quantizer bin of the quantizer Q_{p-1} is split into more number of bins to form the quantizer Q_p. If m is the number of quantizer bins in quantizer Q_{p-1} and each quantizer bin is split into r number of bins then the total number of bins at quantizer Q_p is rm. Figure 4.1 shows an embedded quantizer having three levels ($P = 3$) with $r = 2$.

In order to adopt the concept of embedded quantization in the side quantizers within the MDC scheme to obtain MDSQ-SR a new problem formulation based on the joint decoding distortion constraint is required. For this, P number of quality layers is considered for each of the side quantizer. For the first quantization layer, *i.e.,* $p = 0$, which is also considered as the base layer, the conventional MDSQ design approach discussed in Section 2.5.1 is followed, *i.e.,* deriving the two side quantizers, S_0 and T_0, from the central quantizer C. The rate distortion constraint requirements remains the same as in Eq. (2.3), *i.e.,* the distortion of the combined description decoding is less than the distortion of the individual description decoding at required rate. Then each bin of the side quantizers, S_{p-1} and T_{p-1}, is refined by splitting into r portions leading to refinement side quantizers, S_p and T_p, resulting in the quality enhancement side description layers, s_p and t_p leading to a series of successive refinements of quality of each side description.

As we know, that for any MDC scheme the distortion of the combined description has to be less than the distortion of the individual description at a required rate. Similarly, for the successive refinement of side quantizer-based MDC scheme, the

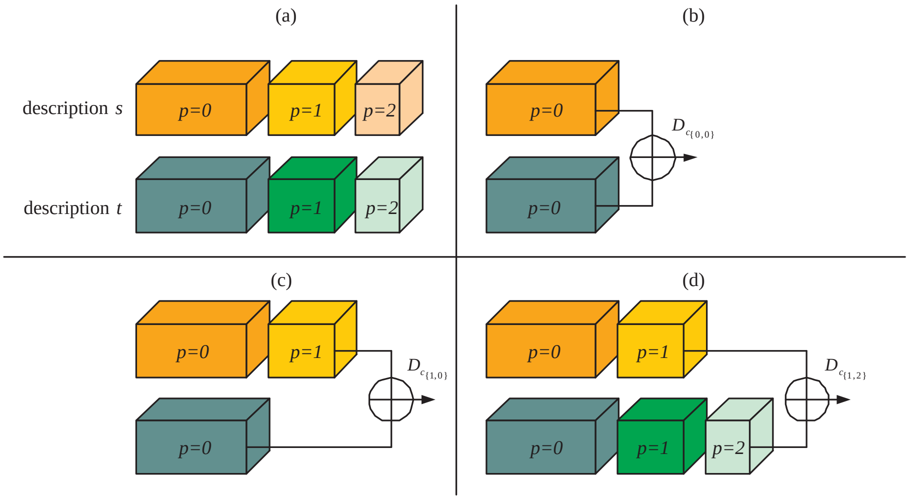


Figure 4.2: Two scalable descriptions from MDSQ-SR having $P = 3$ and three different joint decoding scenarios.

distortion of a side description refined up to any quantization level, p, should be less than that for the side description refined up to the quantization level, $p-1$. For joint decoding two scenarios are considered. Firstly, the joint decoding of side descriptions that are refined up to the same quantization level p. Secondly, the joint decoding of side descriptions that have been refined to unequal quantization levels, p and q, where $p < q$. Such different combinations of joint decoding are shown in Figure 4.2.

In both cases, the distortion of the joint decoding should be less than the individual decoding of the side descriptions. These distortion constraints can be written mathematically as,

$$D_{s_p} < D_{s_{p-1}}, \tag{4.1}$$

$$D_{t_p} < D_{t_{p-1}}, \tag{4.2}$$

$$D_{s_p}, D_{t_p} < D_{c_{p-1}}, \tag{4.3}$$

$$D_{c_p} < D_{s_p}, D_{t_p}, \tag{4.4}$$

$$D_{c_p} < D_{c_{p-1}}, \tag{4.5}$$

provided that $R_{c_p} > R_{c_{p-1}}$, where $p = 0, 1, 2, \cdots, P-1$ and D_{cp} represents the joint decoding distortion of the side descriptions s_p and t_p refined up to level p.

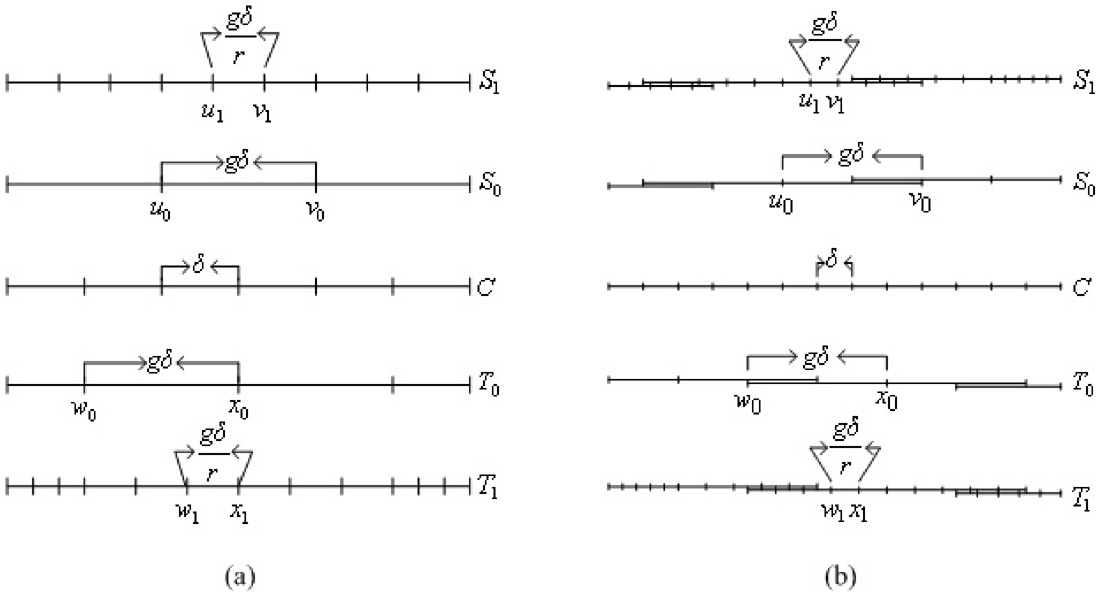


Figure 4.3: Example of MDSQ-SR for $P = 2$ (a) non-overlapped side quantizer bins $r = 3$ (b) overlapped side quantizer bins $r = 5$.

The distortion of a description created from MDSQ is related to that of the other depending on the factor g, defined in Eq. (2.11) and Eq. (2.12). For the refinement layers, the refinement factor, r, also contributes to that relationship. The resulting distortion is due to the quantization bin sizes used in the quantizers. Two examples of a central and corresponding side quantizers for a single level of successive refinement *i.e.,* $P = 2$ for the non-overlapped and overlapped side quantizer scenarios is shown in Figure 4.3. As seen from Figure 4.3, to satisfy the rate-distortion constraints in Eq. (4.1) - Eq. (4.5) for scalable multiple description coding (SMDC) using MDSQ-SR, the relationship of g and r for the separate and joint decoding scenarios of the successively refined side quantizers needs to be established.

4.2.2 Design Conditions

For the central quantizer with a bin size δ, the bin size for any side quantizer at any level p can be written as $\frac{g\delta}{r^p}$, where r is the quantization bin split factor for refinement side quantizers. The bin width factor g of the side quantizer at level $p = 0$ is as shown in Eq. (2.11) and Eq. (2.12) for the non-overlapped and overlapped side quantizer bin merging methods, respectively. Let u_p and v_p be the minimum and the maximum value of a quantizer bin of the side quantizer S_p, and w_p and x_p be the corresponding values for T_p. For any given layer p these values can be written as

$$\begin{aligned} v_p &= u_p + \frac{g\delta}{r^p}, && (4.6) \\ x_p &= w_p + \frac{g\delta}{r^p}. && (4.7) \end{aligned}$$

For two successive levels, p and $p-1$, of the side quantizers S_p the minimum value of the corresponding side quantizer bin can be written as

$$u_p = u_{p-1} + \frac{i_p g\delta}{r^p}, \tag{4.8}$$

where $i_p \in \{0, \cdots, r-1\}$. Now considering the base layer and all successive side quantizer refinements,

$$u_p = u_0 + I_p \frac{g\delta}{r^p}, \tag{4.9}$$

where $I_p = \sum_{\alpha=1}^{p} r^{p-\alpha} i_\alpha$. Note that $I_0 = 0$, as it relates to the base layer. Then,

$$v_p = u_0 + (I_p + 1)\frac{g\delta}{r^p}. \tag{4.10}$$

Similarly, for the side quantizers T_p, the minimum and maximum quantizer bin value is written as,

$$\begin{aligned} w_p &= w_0 + J_p \frac{g\delta}{r^p}, && (4.11) \\ x_p &= w_0 + (J_p + 1)\frac{g\delta}{r^p}, && (4.12) \end{aligned}$$

where $J_p = \sum_{\alpha=1}^{p} r^{p-\alpha} j_\alpha$. with $J_0 = 0$ and $j_\alpha \in \{0, \cdots, r-1\}$. Since two corresponding side quantizer bins in S_p and T_p are either leading or lagging each other with respect to the central quantizer bin, w_0 and u_0 are related as

$$w_0 = u_0 + h\delta, \tag{4.13}$$

where $h \in \{\pm 1, \cdots, \pm(g-1)\}$. Now w_p in Eq. (4.11) and x_p in Eq. (4.12) can be rewritten as,

$$\begin{aligned} w_p &= u_0 + h\delta + J_p \frac{g\delta}{r^p}, && (4.14) \\ x_p &= u_0 + h\delta + (J_p + 1)\frac{g\delta}{r^p}, && (4.15) \end{aligned}$$

Then the bin size Δ_p of the implicit central quantizer for the joint decoding of the refined side quantizers at any level p is formulated as follows:

$$\begin{aligned} \Delta_p &= \min(v_p, x_p) - \max(u_p, w_p), \\ &= \min(u_p + \frac{g\delta}{r^p}, w_p + \frac{g\delta}{r^p}) - \max(u_p, w_p), \\ &= \frac{g\delta}{r^p} + \min(u_p, w_p) - \max(u_p, w_p), \\ &= \frac{g\delta}{r^p} - |u_p - w_p|. && (4.16) \end{aligned}$$

Considering Eq. (4.14) and Eq. (4.15),

$$u_p - w_p = -h\delta - (J_p - I_p)\frac{g\delta}{r^p}. \tag{4.17}$$

The polarity of $J_p - I_p$ corresponds to the leading or lagging relationship of the quantizer bins in S_p and T_p and as the same as that of h, which is the lead/lag factor for the base layer side quantizer bins with respect to the central quantizer bin. Therefore, $|u_p - w_p| = h\delta + (J_p - I_p)\frac{g\delta}{r^p}$ and

$$\Delta_p = \frac{g\delta}{r^p} - h\delta - (J_p - I_p)\frac{g\delta}{r^p}. \tag{4.18}$$

Firstly the separate refinement of individual descriptions is considered and the design conditions for meeting the distortion constraints requirements in Eq. (4.1) - Eq. (4.3) is formulated as follows.

Proposition 4 *The sufficient and necessary condition for satisfying the side quantizer distortion conditions for successive refinements when individual descriptions are independently decoded at level* $p > 0$ *is* $r > 1$*, where* r *is the quantization bin split factor.*

Proof : With reference to Figure 4.3 and for satisfying the successive refinement conditions, Eq. (4.1) and Eq. (4.2), the quantization bin size for the refinement side quantizer at level p must be smaller than that at the previous level, *i.e.,*

$$\frac{g\delta}{r^p} < \frac{g\delta}{r^{p-1}},$$

which is simplified to $r > 1$, since $g\delta > 0$. Therefore, the design condition is $r > 1$ for satisfying distortion constraints $D_{s_p} < D_{s_{p-1}}$ and $D_{t_p} < D_{t_{p-1}}$.

$\blacksquare$.

Proposition 5 *The design condition, $r > g$, where r is the quantization bin split factor and g is the initial bin width factor, is necessary and sufficient for the constraints on the distortion of a refined side description with respect to the distortion due to joint decoding of the side descriptions in the previous level,* i.e., *the constraint Eq. (4.3).*

Proof : For satisfying Eq. (4.3), the quantization bin size for the refinement side quantizer at level p, *i.e.,* $(\frac{g\delta}{r^p})$, must be smaller than the implicit quantizer bin size, Δ_{p-1} when the side quantizers up to the previous level are jointly decoded. This means

$$\begin{aligned} \frac{g\delta}{r^p} &< \Delta_{p-1}, \\ &< \frac{g\delta}{r^{p-1}} - h\delta - (J_{p-1} - I_{p-1})\frac{g\delta}{r^{p-1}}, \\ 1 &< r - h\frac{r^p}{g} - r(J_{p-1} - I_{p-1}). \end{aligned} \tag{4.19}$$

Using mathematical induction, we show that Eq. (4.19) is true for any level of refinement p. For $p = 1$, since $J_{p-1} = I_{p-1} = 0$, Eq. (4.19) becomes

$$\begin{aligned} 0 &< r - 1 - h\frac{r}{g}, \\ r &> \frac{g}{g-h}. \end{aligned} \tag{4.20}$$

For $h_{\max} = g - 1$ and $h_{\min} = 1$ Eq. (4.20) becomes $r > g$ and $r > 1$, respectively. That means Eq. (4.19) is true for $p = 1$, if $r > g$.

Now for level $p = \rho$, let Eq. (4.19) be true, *i.e.*,

$$0 < r - 1 - h\frac{r^{\rho}}{g} - r(J_{\rho-1} - I_{\rho-1}). \tag{4.21}$$

For the following level, $p = \rho + 1$, we can write

$$0 < r - 1 - h\frac{r^{\rho+1}}{g} - r(J_{\rho} - I_{\rho}). \tag{4.22}$$

From the definitions of I_p and J_p, it can be shown that $I_p = rI_{p-1} + i_p$ and $J_p = rJ_{p-1} + j_p$. Using these conditions Eq. (4.22) becomes

$$\begin{aligned} 0 &< r - 1 - r\left[h\frac{r^{\rho}}{g} + (J_{\rho} - I_{\rho})\right], \\ &< r - 1 - r\left[h\frac{r^{\rho}}{g} + r(J_{\rho-1} - I_{\rho-1}) + j_{\rho} - i_{\rho}\right]. \end{aligned} \tag{4.23}$$

From Eq. (4.21),

$$-h\frac{r^{\rho}}{g} - r(J_{\rho-1} - I_{\rho-1}) > 1 - r. \tag{4.24}$$

Also $(j_{\rho} - i_{\rho}) > -r$. Now Eq. (4.23) can be rewritten as

$$\begin{aligned} & r - 1 - r\left[h\frac{r^{\rho}}{g} + r(J_{\rho-1} - I_{\rho-1}) + j_{\rho} - i_{\rho}\right] \\ & \quad > r - 1 - r(1 - r - r), \\ & \quad > 2r^2 - 1. \end{aligned}$$

Since $r > 1$, $(2r^2 - 1) > 0$. Therefore, the condition for $p = \rho + 1$ in Eq. (4.22) becomes true. Therefore, the design conditions are $r > 1$ and $r > g$ for satisfying distortion constraint $D_{s_p}, D_{t_p} < D_{c_{p-1}}$.

■.

Now the joint decoding of refined individual descriptions is considered and the design conditions for meeting the distortion constraints in Eq. (4.4) and Eq. (4.5) is formulated.

Proposition 6 *The design condition, $r^p \neq zg$, where z is a positive integer, r is the quantization bin split factor and g is the initial bin width factor, is necessary* *and sufficient for the constraint on the distortion of joint decoding of refined side descriptions with respect to the distortion due to independent decoding of a refined side description,* i.e., *the constraint in Eq. (4.4).*

Proof: For the level $p-1$, the side quantizer bin size is $\frac{g\delta}{r^{p-1}}$ and the implicit central quantizer bin width for joint decoding is Δ_{p-1}, the limits of which are defined by $[\max(u_{p-1}, w_{p-1}), \min(v_{p-1}, x_{p-1})]$. If the level $p-1$ is refined by a factor r, the refined side quantizer bin width is $\frac{g\delta}{r^p}$. In order to satisfy joint decoding of refined side descriptions (as in Eq. (4.4)), the two side quantizers must overlap, *i.e.,* , $u_p \neq w_p$. This is possible only if the implicit central quantizer bin width for joint decoding of side quantizers at level $p-1$ is not an integer multiplication of the bin size of the side quantizer refined to the next level p. Let z be a positive integer. Then to satisfy Eq. (4.4),

$$\begin{aligned} \Delta_{p-1} &\neq z\frac{g\delta}{r^p}, \\ \frac{g\delta}{r^{p-1}} - h\delta - (J_{p-1} - I_{p-1})\frac{g\delta}{r^{p-1}} &\neq z\frac{g\delta}{r^p}, \\ r - h\frac{r^p}{g} - r(J_{p-1} - I_{p-1}) &\neq z, \end{aligned} \tag{4.25}$$

All variables r, h, p, g, J_{p-1} and I_{p-1} are integers. Therefore, to satisfy the inequality in Eq. (4.25), $\frac{r^p}{g}$ must not be an integer. That means $r^p \neq zg$ must satisfy. Therefore, the design condition is $r^p \neq zg$ for satisfying distortion constraint $D_{c_p} < D_{s_p}, D_{t_p}$.

■.

It should be noted that if constraints in Eq. (4.3) and Eq. (4.4) are satisfied, then the constraint in Eq. (4.5) is also satisfied. Thereby, the MDSQ-SR design conditions are summarized as $r > 1$, $r > g$ and $r^p \neq zg$.

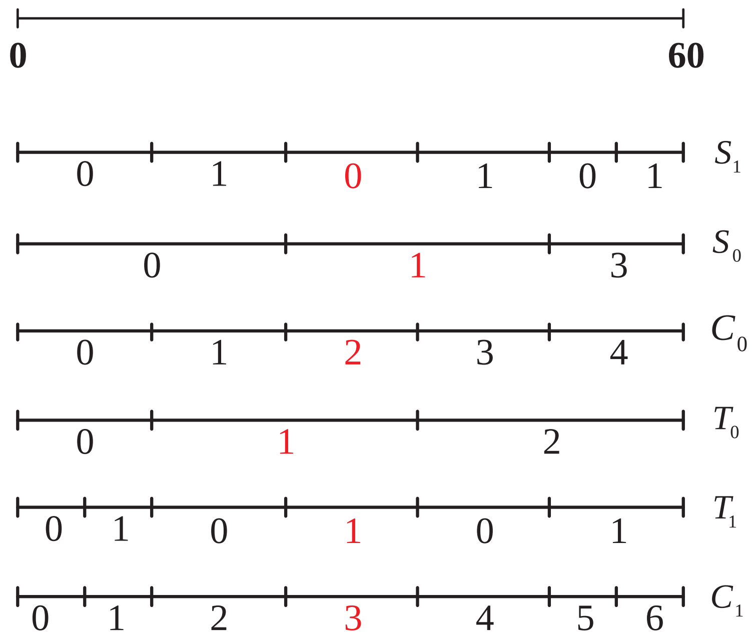
4.2.3 Overlapping Quantization Bin Case Example

Figure 4.4 shows two cases of the side and central quantizers of the MDSQ-SR for two refinement levels *i.e.,* $P = 2$. In Figure 4.4 (a) and (b) C_0, S_0 and T_0

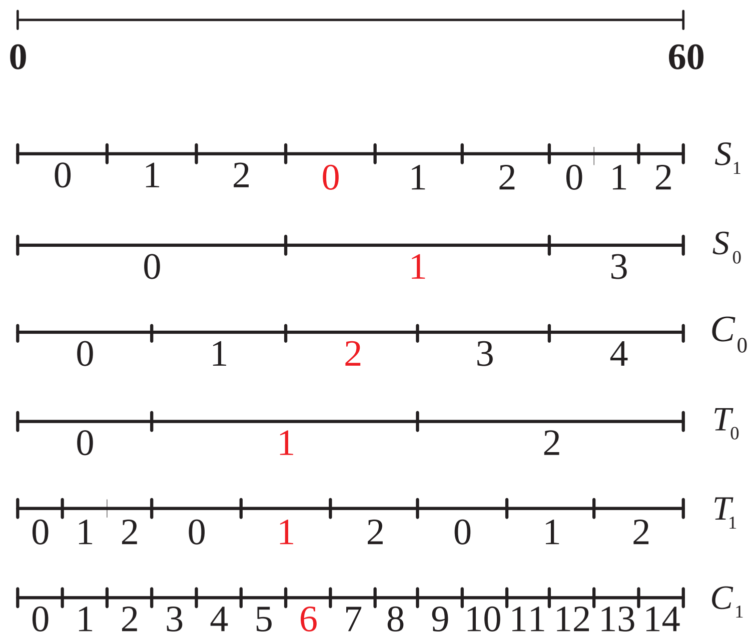
represents the central and the two corresponding side quantizers of the base layer of MDSQ-SR. Similarly S_1 and T_1 represents the side quantizers after refining the bins of the side quantizers S_0 and T_0 respectively. C_1 represents the corresponding central quantizer at level $p = 1$. For both cases shown in Figure 4.4 have the same central and side quantizers for the base layer with parameter $g = 2$. For successive refinement purpose, $r = 2$ and $r = 3$ is used in Figure 4.4 (a) and Figure 4.4 (b) respectively.

The minimum and maximum value of all the quantizers considered in Figure 4.4 is 0 and 60. Therefore, the central quantizer bin width for C_0 is $\Delta_0 = 12$. As $g = 2$ is used the maximum side quantizer bin width for S_0 and T_0 is 24. Now consider a coefficient value of 27 that needs to be quantized using the MDSQ-SR shown in Figure 4.4 (a) and (b). In both cases, the coefficient value of 27 corresponds to the side quantizer bin pairs as (S_0,T_0) = (1,1). For case shown in Figure 4.4 (a) the corresponding side quantizer bins of the side quantizer (S_1,T_1) are (0,1). The corresponding central quantizer bin at level $p = 1$ is 3 whose bin width is same as the bin width of the side quantizer S_1 and S_2. Therefore, there is no change in the dequantization value when both the descriptions are received up to the refinement level $p = 1$ compared to the single description dequantization value.

For the second case shown in Figure 4.4 (b) the corresponding side quantizer bins of the side quantizer (S_1,T_1) are (0,1). The corresponding central quantizer bin at level $p = 1$ is 6 whose bin width is smaller than the bin width of the side quantizers S_1 and S_2. As the bin width is reduced so the distortion is decreased when both the descriptions are received up to the refinement level $p = 1$. From the two cases shown in Figure 4.4 (a) and (b) it can easily be observed that the central quantizer bin width for joint decoding of side quantizers at level $p = 1$ is less than the bin width of the side quantizer al level $p = 1$ and central quantizer at $p = 0$, if the refinement factor r and the side quantizer bin spread g are not integer multiple of each other.



(a)



(b)

Figure 4.4: Overlapping quantization bin case example (a) $g = 2$ and $r = 2$ (b) $g = 2$ and $r = 3$.

Table 4.1: Distortion improvement when joint decoded with respect to the distortion (in dB) of the successively refined side descriptions for 3 refinement layers (For $f = 2$ non-overlapped balanced case).

Images	Refinement factor (r)	$p = 1$	$p = 2$	$p = 3$
Barbara (on chair)	2	0.08	0.47	0.49
Barbara (on chair)	3	6.25	3.83	0.79
Barbara (on chair)	4	0.46	1.51	0.059
Barbara (on floor)	2	0.002	0.002	0.001
Barbara (on floor)	3	6.62	4.31	1.04
Barbara (on floor)	4	0.006	0.006	0.070
Gold Hill	2	0.15	1.33	2.24
Gold Hill	3	6.01	3.24	1.33
Gold Hill	4	1.33	2.78	0.51
Blackboard	2	1.31	1.66	1.72
Blackboard	3	5.1	0.17	0.34
Blackboard	4	1.67	1.70	0.21

4.3 Simulation Results

In this section, the simulation results are presented in four stages. Firstly, the proposed conditions for the MDSQ-SR to fulfil distortion constraints are verified, then the rate distortion performance analysis under lossless channel condition and performance evaluation under packet erasure channel of the MDSQ-SR based image coding are presented. Finally, the computational complexity in terms of execution time of MDSQ-SR based image coding is compared with SDC and EMDSQ based image coding schemes.

4.3.1 MDSQ-SR Designed Condition Verification

In the first set of simulation results, the design conditions proposed in Section 4.2.1 for the MDSQ-SR design are verified. Firstly, non-overlapped balanced side quantizer bins using the staggered index assignment as the base layer with parameter $f = 2$ is considered. In this case $g = 2$ (according to Eq. (2.11)). Table 4.1 shows the improvement of distortion, in terms of PSNR difference (in

Table 4.2: Distortion improvement when joint decoded with respect to the distortion (in dB) of the successively refined side descriptions for 3 refinement layers (For $f = 3$ overlapped balanced case).

Images	Refinement factor (r)	$p = 1$	$p = 2$	$p = 3$
Barbara (on chair)	4	0.17	0.03	0.02
Barbara (on chair)	5	3.5	2.5	0.3
Barbara (on floor)	4	0.05	0.02	0.01
Barbara (on floor)	5	4.0	2.6	0.8
Gold Hill	4	0.16	0.8	0.9
Gold Hill	5	3.5	1.4	0.5
Blackboard	4	0.4	0.8	0.3
Blackboard	5	2.9	2.0	0.9

Table 4.3: Distortion improvements in dBs for the joint decoding of side descriptions refined up to different refinement levels for *Gold Hill* and *Barbara* images for $f = 2$, $r = 3$.

<table>
<tr><th>Images</th><th></th><th colspan="5">Balanced</th></tr>
<tr><td rowspan="6">Barbara (on chair)</td><td rowspan="6">t_p</td><td colspan="5">s_p</td></tr>
<tr><td>p</td><td>0</td><td>0, 1</td><td>0, 1, 2</td><td>0, 1, 2, 3</td></tr>
<tr><td>0</td><td>32.33</td><td>32.39</td><td>39.97</td><td>45.32</td></tr>
<tr><td>0, 1</td><td>32.60</td><td>38.14</td><td>41.39</td><td>45.42</td></tr>
<tr><td>0, 1, 2</td><td>39.84</td><td>40.46</td><td>43.75</td><td>45.72</td></tr>
<tr><td>0, 1, 2, 3</td><td>45.50</td><td>45.58</td><td>45.92</td><td>46.28</td></tr>
<tr><td rowspan="5">Gold Hill</td><td rowspan="5">t_p</td><td>p</td><td>0</td><td>0, 1</td><td>0, 1, 2</td><td>0, 1, 2, 3</td></tr>
<tr><td>0</td><td>31.76</td><td>32.05</td><td>36.11</td><td>44.12</td></tr>
<tr><td>0, 1</td><td>32.46</td><td>36.49</td><td>38.41</td><td>44.18</td></tr>
<tr><td>0, 1, 2</td><td>38.41</td><td>38.65</td><td>41.62</td><td>44.76</td></tr>
<tr><td>0, 1, 2, 3</td><td>44.54</td><td>44.55</td><td>45.31</td><td>45.85</td></tr>
</table>

dB), for joint decoding of refined side descriptions with respect to separate decoding of the individual side descriptions for 3 levels of refinements ($p = 1, ..., 3$) and three different refinement factors ($r \in \{2, 3, 4\}$).

Secondly, the overlapped balanced side quantizer bins using the modified nested index assignment case with parameter $f = 3$ is considered as the base layer, where $g = 4$ (according to Eq. (2.12)). Table 4.2 shows the improvement of distortion, in terms of PSNR difference (in dB), for joint decoding of refined side descriptions with respect to separate decoding of the individual side descriptions for 3 levels of refinements ($p = 1, ..., 3$) and two different refinement factors ($r \in \{4, 5\}$). In both cases, it can be seen that the PSNR improvement is marginal when $r = zg$, where $z = 1$ and $z = 2$, *i.e.,* the case of z being a positive integer, as opposed to when $r \neq zg$ (as in $r = 3$ in Table 4.1 and $r = 5$ in Table 4.2). These results verify the design constraints proposed in Section 4.2.1.

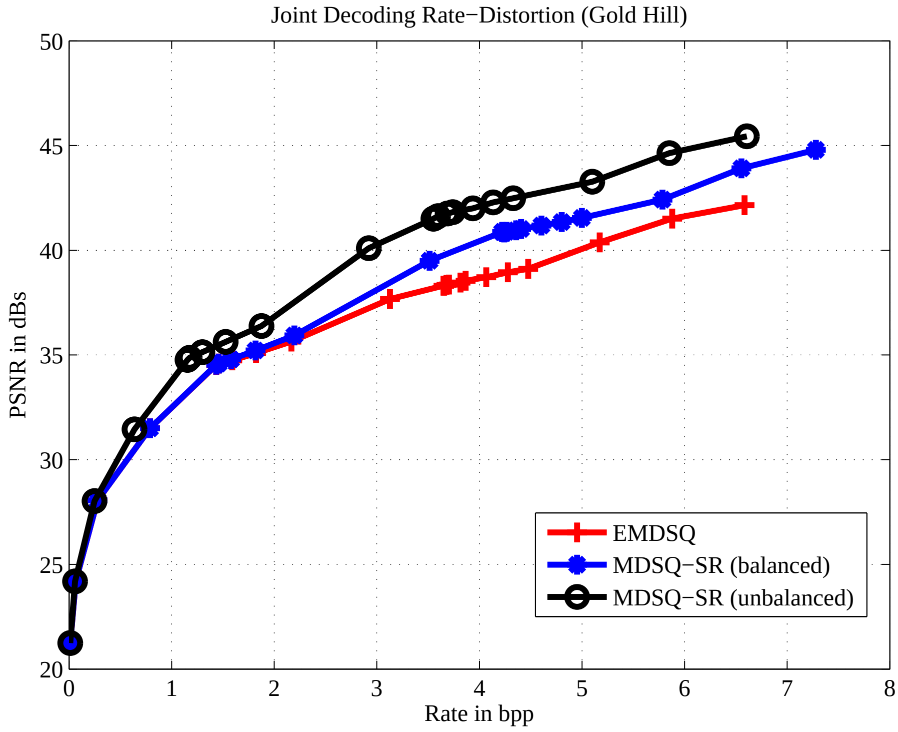
4.3.2 MDSQ-SR Under Lossless Channel Conditions

In the second set of simulation results, the distortion improvement for joint decoding of side descriptions that have been refined to different levels considering the example decoding scenarios shown in Figure 4.2 is demonstrated. Table 4.3 shows the distortion (in PSNR) for 16 different joint decoding scenarios for $P = 4$ MDSQ-SR for two images, *Gold Hill* and *Barbara*. Table 4.3 represent refinement of s and t side descriptions with quantizers S_p and T_p, respectively. The PSNR value in Table 4.3 for $s_p = 0$, $t_p = 0$ corresponds to the joint decoding of the base layer side descriptions from S_0 and T_0. As suggested by the conditions in Eq. (4.1) - Eq. (4.5), Table 4.3 verify the progressive quality refinement of the proposed MDSQ-SR for both example cases shown. In the final set of simulations under lossless conditions, the performance of the proposed MDSQ-SR is compared with that of EMDSQ [53]. Figure 4.5 and Figure 4.6 shows the comparison of the rate-distortion plots for joint and separate decoding of the MDSQ-SR and EMDSQ schemes for the *Gold Hill* and *Barbara (on chair)* images. Similarly, Figure B.6 and Figure B.7 shows the comparison of the rate-distortion plots for joint and separate decoding of the MDSQ-SR and EMDSQ schemes for the *Barbara (on floor)* and *Boats* images. For all the plots $P = 4$ are used and the same base layer central quantizer is used for both MDSQ-SR and EMDSQ schemes. For MDSQ-SR, the non-overlapped index assignment matrix is used where $f = 2$ and $r = 3$. For the joint decoding plots shown in Figure 4.5 and Figure B.7 the effective rate is calculated by adding the rate of two side descriptions together.

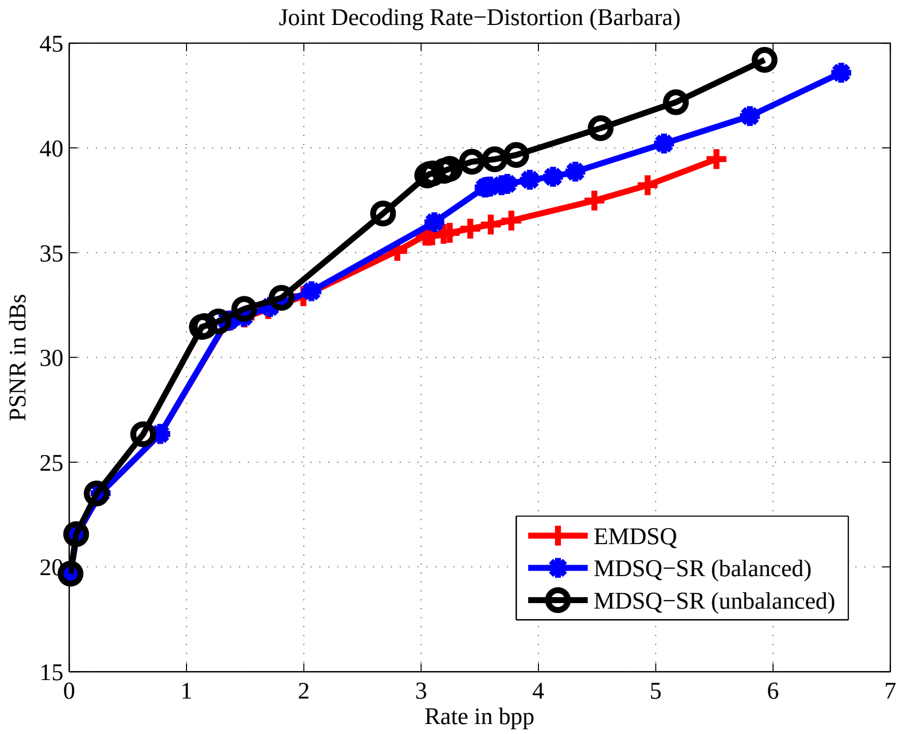
At low data rates, where it corresponds to joint decoding with combination of a side description base layer, all three methods show the same performance. However for more successively refined layers the joint decoding of MDSQ-SR results in better performance compared to central quantizer rate-distortion of EMDSQ. The proposed MDSQ-SR scheme for images when jointly decoded in an unbalanced manner shows PSNR improvement of (1.95 dB for *Gold Hill*, 1.79 dB for *Barbara* on chair, 1.43 dB for *Barbara* on floor, 1.48 dB for *Boats*) with respect to the EMDSQ based MDC scheme for images. Similarly, the proposed MDSQ-SR scheme for images when jointly decode in balanced manner shows PSNR improvement of (1.43 dB for *Gold Hill*, 1.46 dB for *Barbara* on chair, 1.25 dB for *Barbara* on floor, 1.29 dB for *Boats*) with respect to the EMDSQ based MDC scheme for images. The proposed MDSQ-SR scheme for images when jointly decoded in an unbalanced fashion is on average 0.3 dB better than the balanced joint decoding of the proposed MDSQ-SR scheme for all the test images. Also the MDSQ-SR started with a base layer having unbalanced side descriptions performed better than the MDSQ-SR started with a base layer having balanced side descriptions. It is also clear from results in Figure 4.6 and Figure B.7 that the successive refinement of side quantizers gives better side rate distortion performance. Figure 4.7 and Figure B.8 shows side description performance for MDSQ-SR when started with base layer having unbalanced side descriptions. The refined side descriptions also remain unbalanced as seen from the plots.

4.3.3 MDSQ-SR Under Lossy Channel Conditions

The goal of this set of simulation is to evaluate the performance of the proposed MDSQ-SR under lossy channel conditions. In this case we considered the transmission over a packet erasure channel. For this purpose, each side description layer data is packetized and different packet loss percentages are considered. Let M be the total number of packets for each description and l be the number of

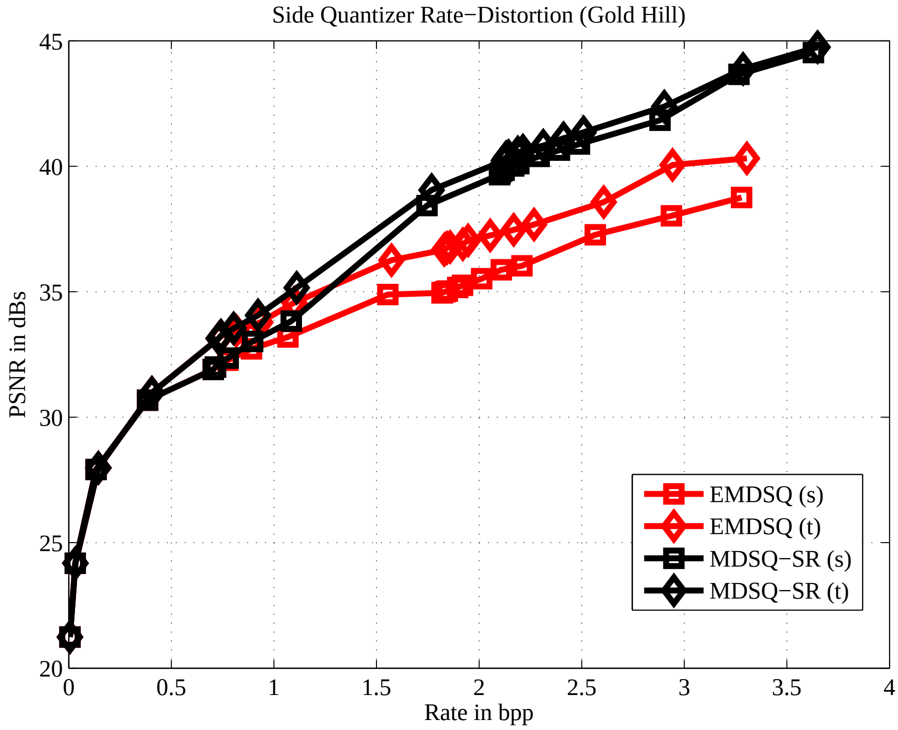


(a)

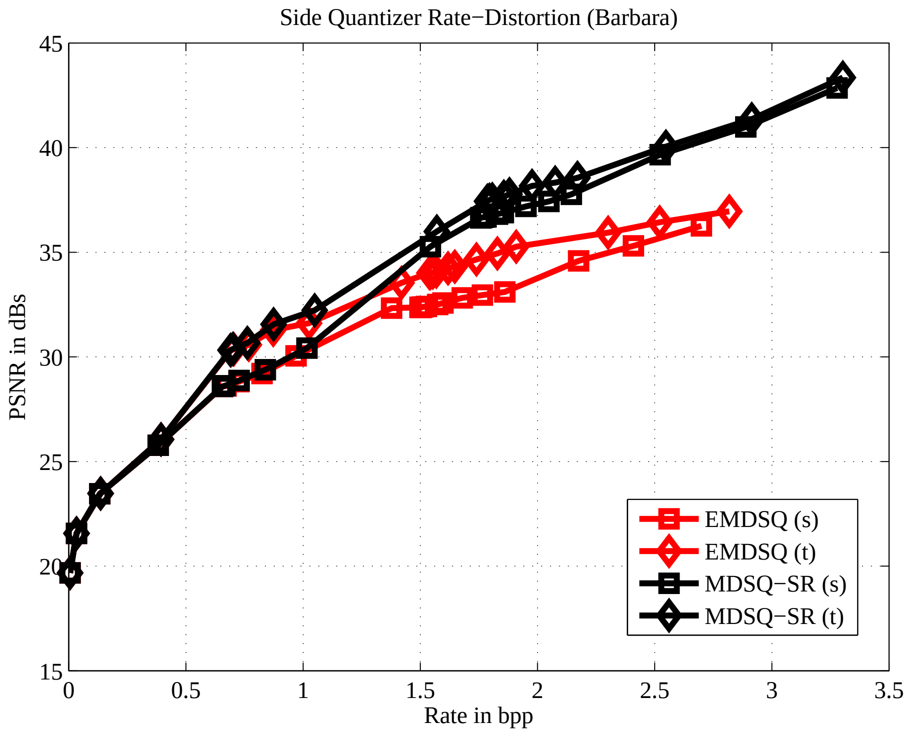


(b)

Figure 4.5: Joint decoding comparison of the proposed MDSQ-SR and EMDSQ schemes for (a) *Gold Hill* (b) *Barbara (on chair)* images.

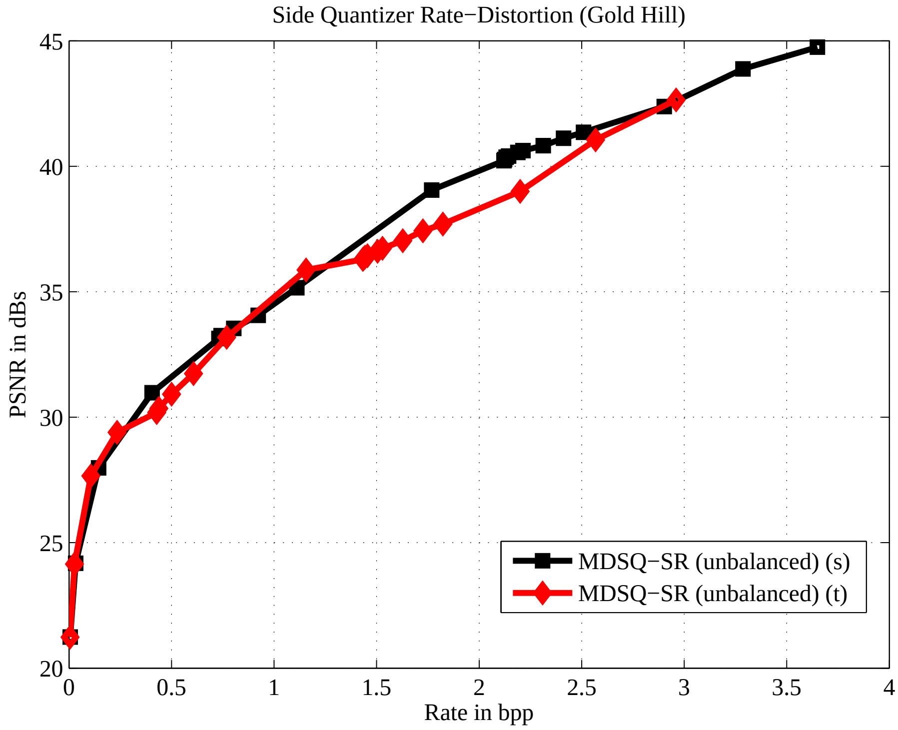


(a)

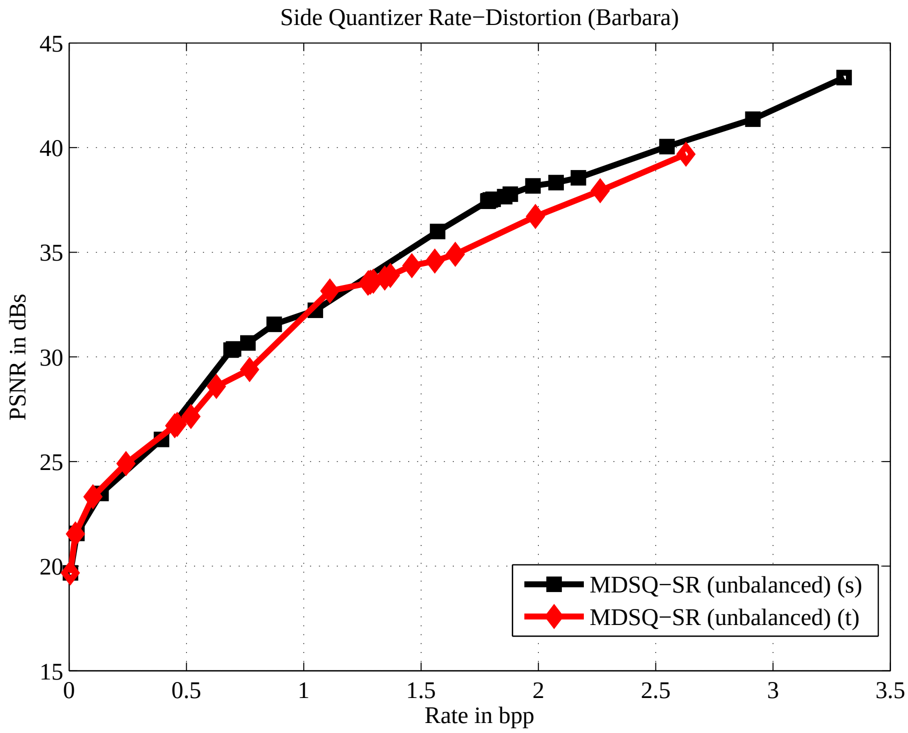


(b)

Figure 4.6: Balanced side description decoding comparison of the proposed MDSQ-SR and EMDSQ schemes for (a) *Gold Hill* (b) *Barbara (on chair)* images.

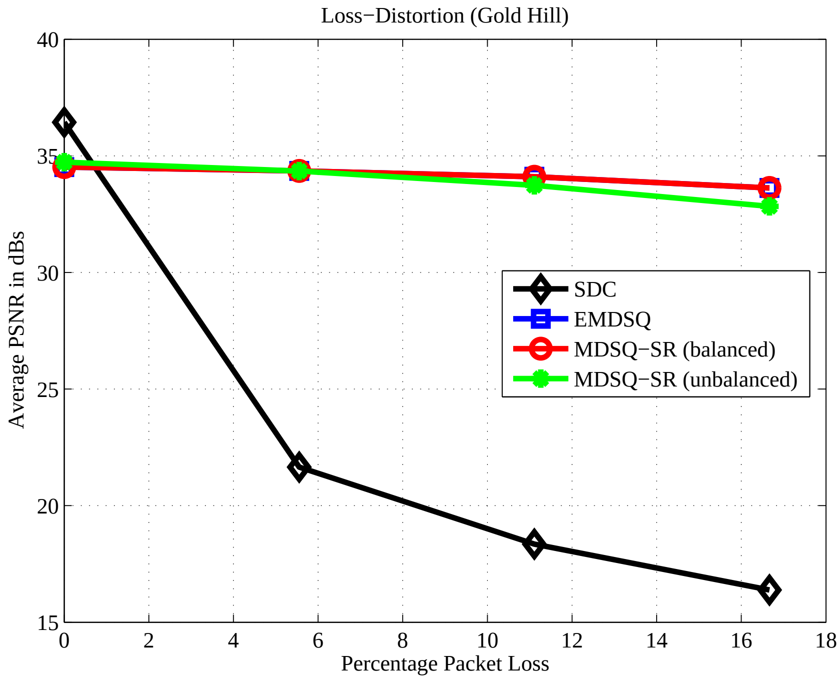


(a)

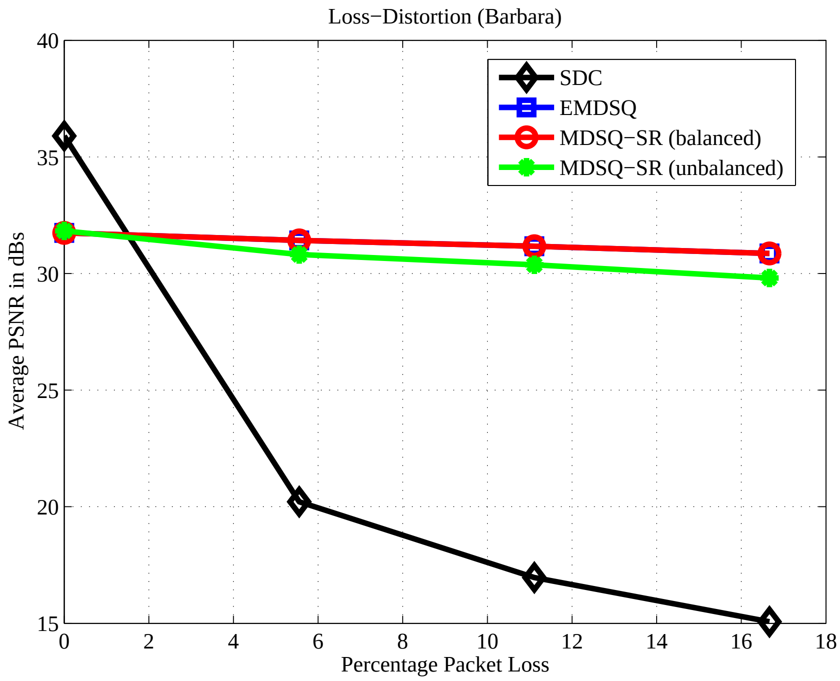


(b)

Figure 4.7: Unbalanced side description decoding of the proposed MDSQ-SR schemes for (a) *Gold Hill* (b) *Barbara (on chair)* images.

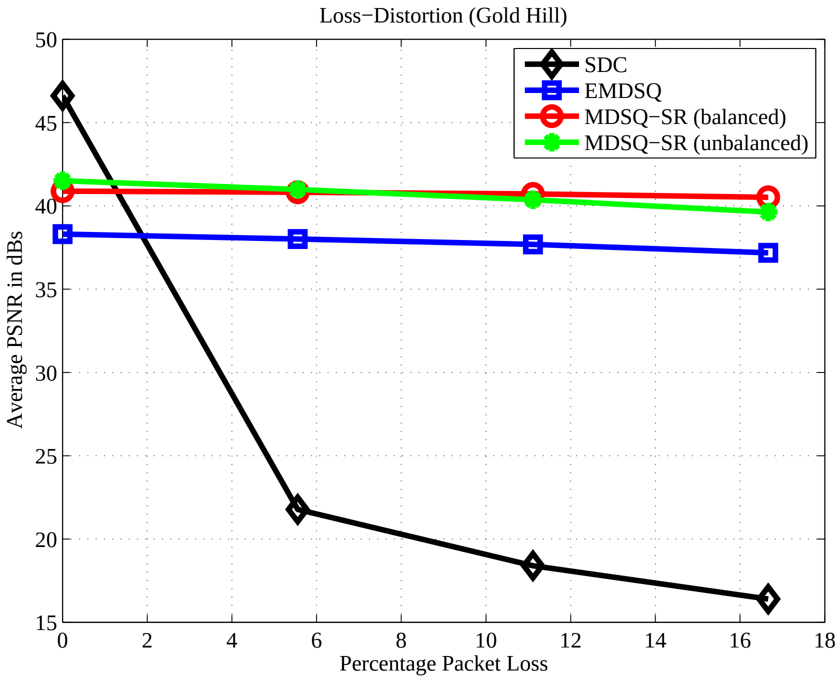


(a)

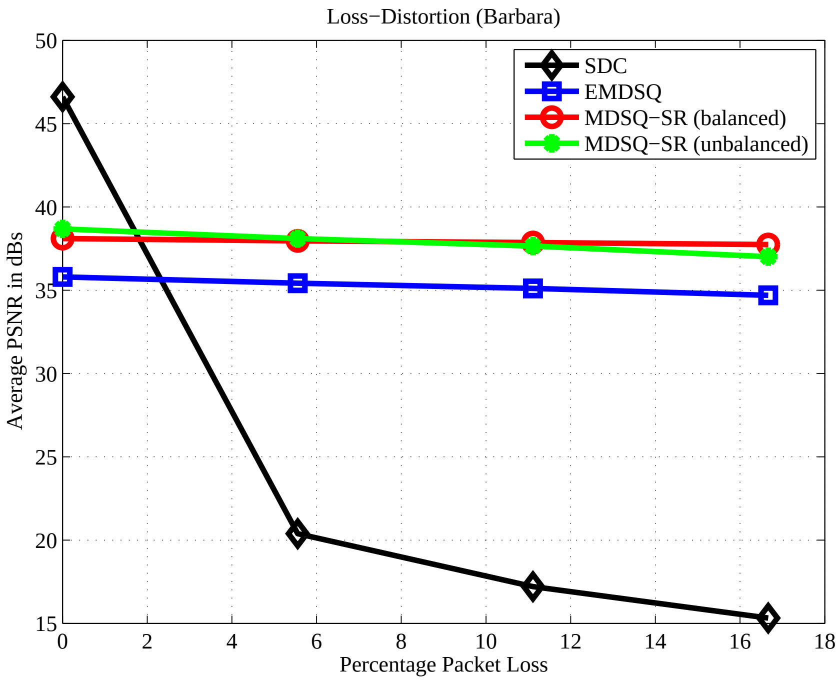


(b)

Figure 4.8: Performance comparison of the proposed MDSQ-SR, EMDSQ and SDC schemes under packet erasure channel for (a) *Gold Hill* (b) *Barbara (on chair)* images encoded at 0.5 bpp/description.



(a)



(b)

Figure 4.9: Performance comparison of the proposed MDSQ-SR, EMDSQ and SDC schemes under packet erasure channel for (a) *Gold Hill* (b) *Barbara (on chair)* images encoded at 2.0 bpp/description.

Figure 4.10: Portion of the decoded images from single description decoding (left) and joint decoding from MDSQ-SR (right) after 18% packet loss.

lost packets. Then there is a total of ${}^{M}C_{l}$ number of combinations to loose l packets from the total of M packets. The average PSNR for a particular packet loss percentage is then calculated by measuring and averaging the PSNR of all possible combinations. For this set of experiments the packet erasure channel model having the loss rate varying between 0% and 18% is considered.

The robustness performance, in terms of average PSNR, of the decoded image for MDSQ-SR, EMDSQ and SDC at different packet loss rates for two different data rates of 0.5 bpp/description and 2.0 bpp/description is shown in Figure 4.8 and Figure 4.9 respectively for *Gold Hill* and *Barbara (on chair)* images. For MDSQ-SR we considered the two cases: non-overlapped $f = 2$ with balanced and unbalanced base layer side descriptions and other parameters being $r = 3$ and $P = 4$. For SDC, conventional wavelet-based progressive coding is used. SDC

SDC ($p = 0$),
14.39 dB

SDC ($p = 0, 1$),
14.39 dB

SDC ($p = 0, 1, 2$),
14.39 dB

Figure 4.11: A part of the decoded *Barbara (on floor)* image from single scalable description after 22% packet loss in each refinement layer.

shows the best performance in the case of no packet loss. This is mainly because rate increment seen in MDC-based solutions, which can be regarded as the cost of making the data streams robust. However, the performance of SDC scheme deteriorates rapidly even for a small percentage of packet loss. As the number of packets lost increases the PSNR of MDSQ-SR based scheme is decreased with less rate as compared to the EMDSQ scheme, for high data rate MDC. However, the performance of both schemes is comparable for low data rates. Furthermore, the balanced description joint decoding case appears to be more robust compared to the unbalanced joint decoding case for low data rates and high packet loss rates. Figure 4.10 shows the portion of the decoded *Gold Hill* and *Barbara (on chair)* images from single description decoding and joint description decoding of the MDSQ-SR, where the encoding rate for both cases are 0.5 bpp. The superior visual quality of the MDSQ-SR based decoding at 0.5 bpp data rate with 18% packet loss rate can easily be observed from Figure 4.10 when compared with single description decoding.

The loss-distortion curves shown in Figure 4.8 and Figure 4.9 considers equal amount of packet loss rate in each description and refinement layer for joint decoding up to the same refinement levels for each description. The effect of equal and different packet loss rate on each description and refinement layer, jointly decoded at different refinement layer is also studied. Figure 4.11 shows the portion of the decoded *Barbara (on floor)* image and PSNR value in dB when decoded from single scalable description after 22% packet loss in each layer for

s_p $(p = 0)$,
t_p $(p = 0)$
30.54 dB

s_p $(p = 0)$,
t_p $(p = 0, 1)$
33.96 dB

s_p $(p = 0)$,
t_p $(p = 0, 1, 2)$
35.16 dB

s_p $(p = 0, 1)$,
t_p $(p = 0)$
33.65 dB

s_p $(p = 0, 1)$,
t_p $(p = 0, 1)$
38.05 dB

s_p $(p = 0, 1)$,
t_p $(p = 0, 1, 2)$
41.38 dB

s_p $(p = 0, 1, 2)$,
t_p $(p = 0)$
34.68 dB

s_p $(p = 0, 1, 2)$,
t_p $(p = 0, 1)$
40.91 dB

s_p $(p = 0, 1, 2)$,
t_p $(p = 0, 1, 2)$
43.30 dB

Figure 4.12: A part of the decoded *Barbara (on floor)* image for different joint description decoding scenarios after 22% packet loss in each description and refinement layer.

s_p $(p = 0)$,
t_p $(p = 0)$
31.06 dB

s_p $(p = 0)$,
t_p $(p = 0, 1)$
33.97 dB

s_p $(p = 0)$,
t_p $(p = 0, 1, 2)$
35.16 dB

s_p $(p = 0, 1)$,
t_p $(p = 0)$
36.02 dB

s_p $(p = 0, 1)$,
t_p $(p = 0, 1)$
38.30 dB

s_p $(p = 0, 1)$,
t_p $(p = 0, 1, 2)$
41.12 dB

s_p $(p = 0, 1, 2)$,
t_p $(p = 0)$
38.50 dB

s_p $(p = 0, 1, 2)$,
t_p $(p = 0, 1)$
42.20 dB

s_p $(p = 0, 1, 2)$,
t_p $(p = 0, 1, 2)$
43.40 dB

Figure 4.13: A part of the decoded *Barbara (on floor)* image for different joint description decoding scenarios after 22% and 11% packet loss in description s and t.

different refinement layers. It can easily be observed from Figure 4.11 that the decoding quality is not improved in single scalable description case even when more refinement layers are available at the decoder.

Figure 4.12 shows the portion of the decoded *Barbara (on floor)* image and their PSNR values in dB for different joint decoding scenarios after 22% packet loss rate in each description and refinement layer. It is evident from Figure 4.12 that the joint decoding quality improves as more refinement layers are available at the decoder for the proposed MDSQ-SR scheme. Similarly, Figure 4.13 shows the portion of the decoded *Barbara (on floor)* image and their PSNR values in dB for different joint decoding scenarios after 11% and 22% packet loss rate in description s_p and t_p, respectively. It is evident from Figure 4.13 that the joint decoding quality improves as more refinement layers are available at the decoder. Furthermore, the quality of the joint decoding from low packet drop rate description with more refinement layers is better than the joint decoding quality from high packet loss rate description with more refinement layers.

4.3.4 Computational Complexity of MDSQ-SR

In this section, computational complexity of the proposed MDSQ-SR based MDC, EMDSQ based MDC and single scalable description coding is calculated and compared. The criteria used to measure the computational complexity is in terms of execution time as suggested in [108].

Let K be the total execution time required for any image coding system. Then for single wavelet based image coding system, the execution time K_{SDC} can be written in terms of each coding block as

$$K_{SDC} = K_{DWT} + K_{SQ} + K_{EC}, \tag{4.26}$$

where K_{DWT}, K_{SQ} and K_{EC} are the execution time required by the DWT, scalar quantization and entropy coding blocks. Similarly, the execution time for any quantization based MDC system having two descriptions can be written as

$$K_{MDC} = K_{DWT} + K_{MQ} + 2 * K_{EC}, \tag{4.27}$$

Table 4.4: Execution time in seconds for quantization block of SDC.

	Iterations							
Images	1	2	3	4	5	Mean	Variance	VMR
Gold Hill	3.56	3.48	3.48	3.48	3.45	3.49	0.00170	0.00048
Barbara (on chair)	3.53	3.57	3.53	3.57	3.57	3.55	0.00048	0.00013
Barbara (on floor)	3.53	3.53	3.48	3.50	3.48	3.50	0.00063	0.00018
Blackboard	3.45	3.42	3.50	3.51	3.50	3.47	0.00150	0.00044
Boats	3.45	3.50	3.50	3.53	3.51	3.49	0.00087	0.00024

Table 4.5: Execution time in seconds for MDSQ-SR block of SMDC.

	Iterations							
Images	1	2	3	4	5	Mean	Variance	VMR
Gold Hill	4.70	4.81	4.78	4.81	4.95	4.81	0.00810	0.00170
Barbara (on chair)	4.87	4.85	4.87	4.89	4.89	4.87	0.00028	0.00006
Barbara (on floor)	4.76	4.79	4.73	4.82	4.76	4.77	0.00120	0.00024
Blackboard	4.85	5.06	4.75	4.76	4.89	4.86	0.01580	0.00320
Boats	4.95	4.81	4.84	4.85	4.85	4.85	0.00310	0.00064

Table 4.6: Execution time in seconds for EMDSQ block of SMDC.

	Iterations							
Images	1	2	3	4	5	Mean	Variance	VMR
Gold Hill	8.82	8.81	8.68	8.71	8.71	8.74	0.00410	0.00047
Barbara (on chair)	9.14	9.01	9.01	8.95	9.07	9.03	0.00520	0.00057
Barbara (on floor)	8.84	8.79	8.82	8.71	8.79	8.79	0.00240	0.00027
Blackboard	8.84	8.84	8.92	8.89	8.93	8.88	0.00180	0.00025
Boats	8.87	8.85	9.06	8.96	8.82	8.91	0.00960	0.00110

where K_{MQ} is the execution time required for creating two descriptions from quantization block.

So for the comparison of the complexity of SDC and MDC in terms of execution time, the only changed block is the quantization block. Therefore, in this section $K_{MDSQ-SR}$, K_{EMDSQ} and K_{SQ} are calculated and compared. For calculation of the execution time, "cputime" function of the Matlab is used. Wavelet based SDC, MDSQ-SR and EMDSQ based SMDC encoder modules are executed on X86-based PC having 2.2 GHz microprocessor and 2GB of memory. For each coding system, execution time is calculated for 5 times with the same coding configuration. To evaluate the precision of the execution time, variance to mean ratio (VMR) is also calculated. The zero VMR value indicates that all the measured execution time is identical for all iterations. VMR value greater than zero indicates the loss of accuracy in execution time measurement.

Table 4.4, Table 4.5 and Table 4.6, shows the execution time, mean, variance and VMR for five iterations having same configuration for all image data set for SDC, MDSQ-SR based MDC and EMDSQ based MDC respectively. From Table 4.4 and Table 4.5 it can easily be observed that the execution time increased by the MDSQ-SR is 1.33 seconds when compared with the execution time of the scalar quantizer of SDC. Similarly, from Table 4.5 and Table 4.6 it can easily be observed that the execution time of MDSQ-SR is reduced by 4.04 seconds when compared with the EMDSQ based MDC. The proposed MDSQ-SR results into 83% less complex MDC scheme than EMDSQ based MDC scheme but 38% more complex than the SDC scheme.

4.4 Summary

In this chapter, a new framework for SMDC based on MDSQ was discussed. In the proposed MDSQR-SR based SMDC, the base layer of each of the side description was successively refined to provide the quality scalability in both side and central decoding. The base layer of the MDSQ-SR was designed by a well-known MDSQ method and different index assignment matrices were used to change the amount of redundancy between the descriptions. Joint balanced and unbalanced description decoding in terms of number of refinement levels is also provided by the MDSQ-SR.

The relationship between the design parameters, the refinement factor r and the side quantizer spread factor g to meet the distortion constraints of the proposed MDSQ-SR scheme was derived in this chapter. It is proposed that the design conditions, $r > 1$, $r > g$ and $r^p \neq zg$, where z is a positive integer and p is the refinement levels, are jointly sufficient and necessary for satisfying the distortion constraints and verified through simulation results. The performance of the MDSQ-SR for various base layer options is evaluated and compared with EMDSQ and single description coding. The unbalanced and balanced joint decoding from the proposed MDSQ-SR scheme has shown the average PSNR improvement of 1.66 dB and 1.35 dB with respect to the joint decoding from EMDSQ scheme for all the test image data set. In addition to its superior rate-distortion performance than EMDSQ, MDSQ-SR has also simplified the design of generating scalable multiple descriptions and also reduced the computational complexity by 83% because only one index assignment matrix is used at the base layer and the refinement layers are generated by successively refining the side quantizers of the base layer.

Chapter 5

Fully Scalable Multiple Description Image Coding

In the previous chapters, schemes for creating more than two quality scalable descriptions using successive refinement of the side quantizers are presented. In this chapter, a novel scheme that incorporates both the quality and resolution scalability in multiple description image coding framework is presented. Different resolution descriptions are created by using MUDC concept for different wavelet decomposition levels and different quality enhancement layers of each description is generated by the successive refinement of the side quantizers. The robustness of the fully scalable multiple description image coding is further increased by exploiting the correlation information generated from MDSQ. The correlated information from MDSQ-based descriptions, which is called side information and is embedded in each description and can be extracted at the decoder from each individual description. The motivation for such type of encoding is to improve the side decoding and to increase the overall robustness of the scheme. The rest of the chapter is organized as follows: The proposed scheme that incorporates both the quality and resolution scalability is presented in Section 5.1. Section 5.2 focuses on the encoding and decoding procedure of MDSQ-based MDC scheme using side information followed by the summary in Section 5.3.

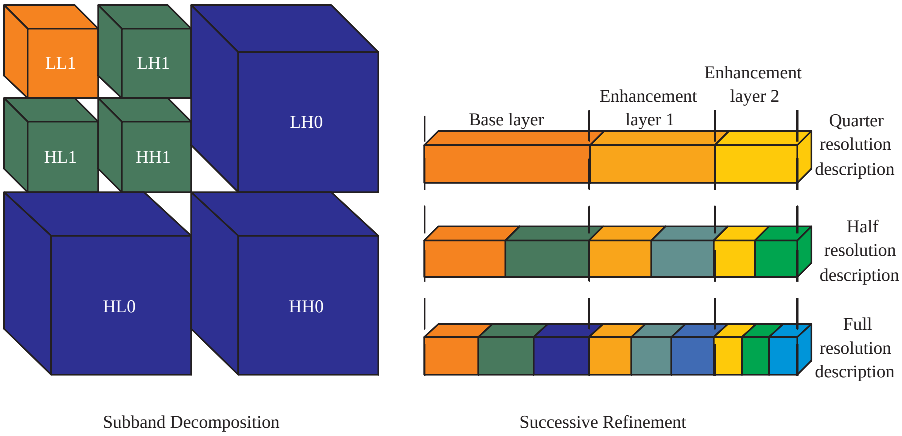


Figure 5.1: Different resolution description creation procedure.

5.1 Multiple Description Image Coding using Multi-channel MDSQ-SR

In this section, a MDC based approach for robust scalable coding that incorporates both the quality and spatial resolution scalability functionalities into the wavelet and MDSQ-based MDC is presented. For achieving quality scalability in each description MDSQ-SR is used, in which the base layer is designed by simple MDSQ and the enhancement layers are generated by continuously refining the side quantizer bins. The theoretical framework of MDSQ-SR is defined in Chapter 4 and it is evident that MDSQ-SR based approach is more efficient in terms of computational complexity, coding performance and robustness as compared to EMDSQ [54]. Different resolution descriptions with different quality scalability parameters are created by considering MUDC. The distortion constraints and conditions of different descriptions selection from several MDSQs are defined in Chapter 3.

For any scalable coding, multi-resolution decomposition and successive refinement of the quantization are the two main components. So for the proposed fully SMDC, multi-resolution decomposition is achieved by the DWT that decomposes the image into different subbands and quality scalability is achieved by successively refining the side quantizers. In general DWT-based MDC framework for two descriptions, all the subbands are considered in each description, so every

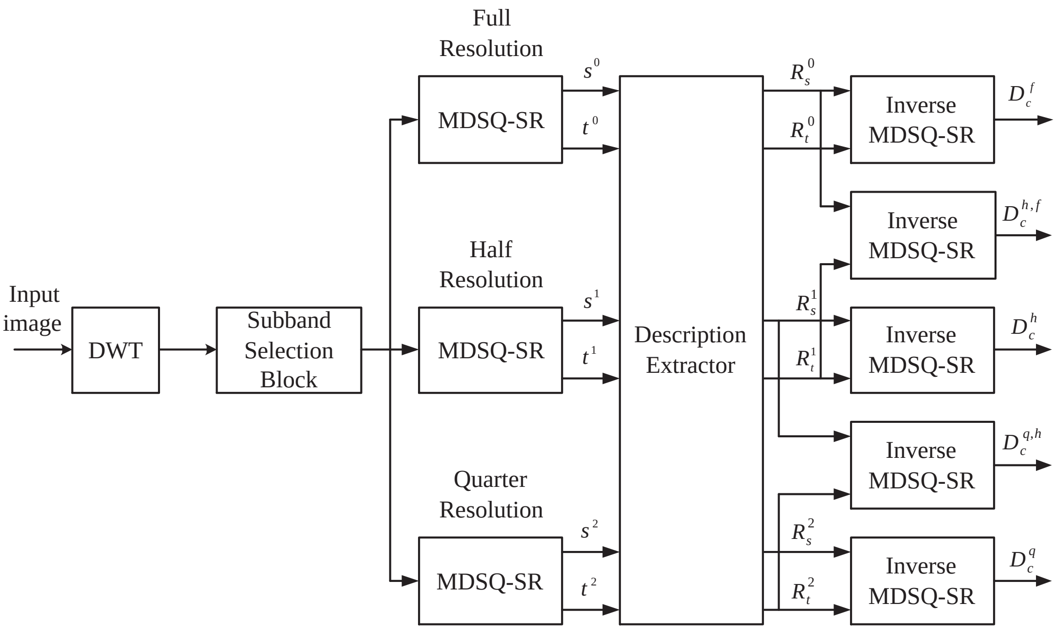


Figure 5.2: Different resolution based multiple description scheme using several MDSQ-SR's.

description is composed of all the coefficients. But in the proposed fully SMDC scheme, different spatial resolution descriptions are generated by selecting the required subbands for a particular decoded resolution. Figure 5.1 shows how the base and enhancement layer for quarter, half and full resolution descriptions are created. The resolution of the decoded image is the same as the highest resolution among all the received descriptions. The block diagram of such a system is shown in Figure 5.2.

In the proposed fully SMDC scheme, firstly the input image is decomposed into different subbands by using the DWT. The subband selection block selects the required subbands for quarter, half and full resolution image decoding. The selected subbands for a particular spatial resolution are then passed to different MDSQ-SRs having different parameters (g, a, r, p) that generate two quality scalable descriptions for quarter, half and full spatial resolution. Let D_c^q, D_c^h, D_c^f be the joint decoding distortion of the quarter, half and full resolution respectively, when both the descriptions from the same spatial resolution MDSQ-SR are received at the decoder. Similarly $D_c^{q,h}$, $D_c^{h,f}$, $D_c^{q,h,f}$ be the joint decoding distortion of the half and full resolution decoding when the descriptions from different spatial resolution MDSQ-SRs are received. The distortion constraints for such kind of decoding are related to each other as

$$D_c^{q,h} \leq D_c^h, \tag{5.1}$$

$$D_c^{h,f} \leq D_c^f, \tag{5.2}$$

$$D_c^{q,h,f} \leq D_c^{h,f}. \tag{5.3}$$

Fewer subbands are required for lower spatial resolution descriptions compared to higher spatial resolution descriptions. Therefore, to satisfy the constraints in Eq. (5.1)-Eq. (5.3) the central quantizer bin size of the MDSQ-SR for lower resolution description is set smaller than the central quantizer bin size of the MDSQ-SR for higher resolution description. In other words, the MDSQ index for the lowest spatial resolution description is greater than the MDSQ index of the higher spatial resolution as done in MUDC. Following the similar principle as in MDSQ-SR, each side quanitzer of MDSQ of different spatial resolution is then successively refined to provide quality scalability in each description. In this way, the generated descriptions can be decoded jointly in unbalanced way not only in terms of spatial resolution but also in terms of MDSQ-SR parameters for different quality layers. The extractor block extracts the descriptions according to the data rate requirement by selecting the subset of subbands and refinement levels of the MDSQ-SR to meet the spatial resolution and quality preferences of the end user in such a way that the joint distortion is minimum.

5.1.1 Rate-Distortion-based Extraction

Once the scalable descriptions s and t at a particular spatial resolution are encoded, the extractor block extracts the descriptions according to the data rate requirement by selecting the subset of spatial subbands and refinement levels of the MDSQ-SR according to the resolution and quality preferences. The extracted description s_e and t_e are adapted in such a way to minimize the joint distortion at required rate. For this purpose, distortion expression is required that considers all the contribution parameters like spatial decomposition subbands and successive refinement levels of the MDSQ-SR. For any discrete signal and its approximation, the distortion is measured by MSE and is calculated by using the Eq. (2.1).

Let I denote the spatial decomposition of the input image and $\hat{I}$ be the reconstructed version of the spatial coefficients after the inverse quantization. The quantization is the only contributing factor to the distortion associated to $\hat{I}$, as the transformation and entropy coding are perfectly invertible. The distortion measured in spatial and frequency domain is same if the spatial transforms are orthonormal. For non-orthonormal transforms, the distortion in transform domain approximates the distortion in spatial domain. For the rate distortion based extraction, the distortion in transform domain is considered to avoid the inverse spatial transformation for each successive refinement. Let G be the total number of spatial decomposition levels of the input image and $\hat{W}_d$ be the subband after spatial decomposition, where $d = 0, \cdots, 3G$, then the distortion of image in transform domain can be written as,

$$D(\hat{I}) = \sum_{d=0}^{3G} D(\hat{W}_d), \tag{5.4}$$

Two scalable descriptions are created by using MDSQ-SR for P refinement levels, then the contribution by each spatial subband for each refinement level p to the total distortion is,

$$D(\hat{I}) = \sum_{d=0}^{3G} \sum_{p=0}^{P-1} D(\hat{W}_{d,p}), \tag{5.5}$$

The distortion considered in Eq. (5.5) is only by the quantization of the spatially decomposed image. Similarly, the rate contribution for each spatial subband and successive refinement to overall rate is,

$$R(I) = \sum_{d=0}^{3G} \sum_{p=0}^{P-1} R_{d,p}, \tag{5.6}$$

Let R_s and R_t be the bit rate requirement of each extracted description, then the extracted description s_e and t_e from the scalable descriptions s and t are adapted in such a way to minimize the joint distortion D_c at required rate $R_c = R_s + R_t$ by selecting different spatial subbands and refinement levels. This can be written

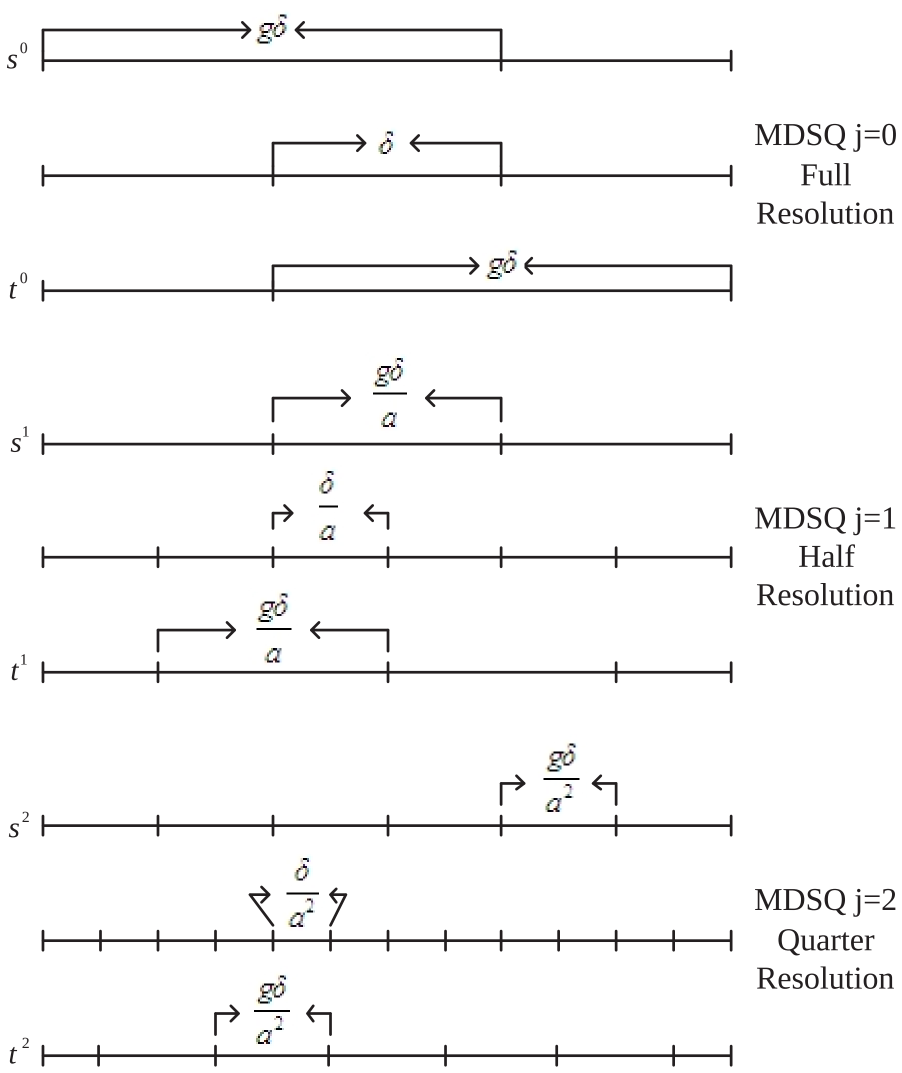


Figure 5.3: Central and side quantizers for different resolution description.

mathematically in Eq. (5.7) as

$$min\ D_c(\hat{I})\ subject\ to\ R_c \leq R_s + R_t. \tag{5.7}$$

The distortion contribution in Eq. (5.5) is considered by selecting all the spatial subbands and different refinement levels p in each description s and t. For different spatial resolution descriptions the required spatial subbands at that resolution are considered from each description s and t *i.e.,* ($d = 0, 1, \cdots, 3G - 6$ for quarter and $d = 0, 1, \cdots, 3G - 3$ for half spatial resolution).

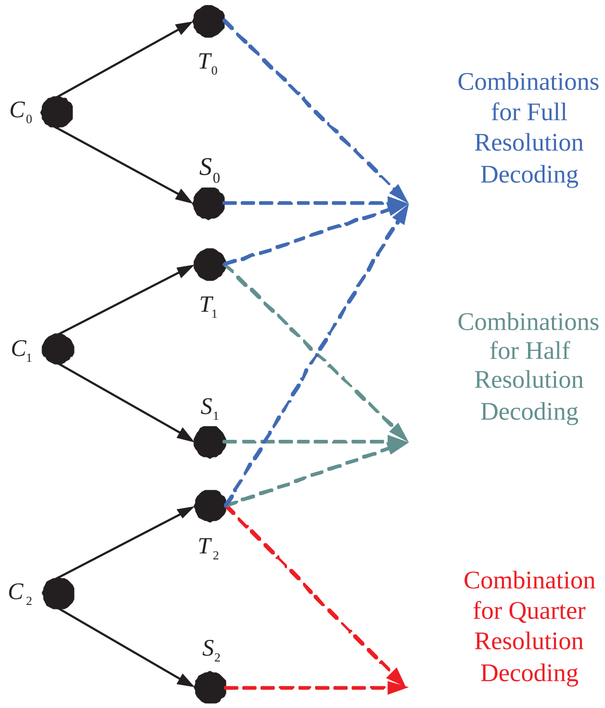


Figure 5.4: Joint decoding at different resolution by combining two descriptions from different MDSQs.

5.1.2 Simulation Setup and Results

For both the resolution and quality scalability, $N = 6$ descriptions using $J = 3$ MDSQs with successive refinement are considered. Let $j = 0, 1, 2$ be the MDSQ index number used for full, half and quarter resolution descriptions respectively. Let D_{sj_p} and D_{tj_p} be the distortion of side description with p number of refinement levels. Then the distortion constraint of individual and joint decoding from each MDSQ-SR remains the same as mentioned in Eq. (4.1) - Eq. (4.5). Similarly the distortion constraints for joint decoding of description from different MDSQ-SR remains same as mentioned in Eq. (3.1) - Eq. (3.8).

Different conditions on the parameters of different MDSQ-SR are proposed for joint description decoding from different MDSQ-SR and from the same MDSQ-SR in Chapter 3 and Chapter 4 respectively. Let δ be the central quantizer bin size of the initial MDSQ which in this case is the central quantizer bin size of full

Table 5.1: Quantizer bin width for the base layer of single and joint decoding description at full resolution from same and different MDSQ-SRs.

Description received	Decoded resolution	Quantizer bin size
s^0	Full	$g\delta$
t^0	Full	$g\delta$
s^1	Half	$\frac{g\delta}{a}$
t^1	Half	$\frac{g\delta}{a}$
s^2	Quarter	$\frac{g\delta}{a^2}$
t^2	Quarter	$\frac{g\delta}{a^2}$
(s^0, t^0)	Full	δ
(s^0, s^1)	Full	$\frac{g\delta}{a}$
(s^0, t^1)	Full	$\frac{\delta}{a}$ or $\frac{g\delta}{a}$
(t^0, s^1)	Full	$\frac{g\delta}{a}$
(t^0, t^1)	Full	$\frac{\delta}{a}$ or $\frac{g\delta}{a}$
(s^0, s^2)	Full	$\frac{g\delta}{a^2}$
(s^0, t^2)	Full	$\frac{\delta}{a^2}$ or $\frac{g\delta}{a^2}$
(t^0, s^2)	Full	$\frac{g\delta}{a^2}$
(t^0, t^2)	Full	$\frac{\delta}{a^2}$ or $\frac{g\delta}{a^2}$

description MDSQ and is related to other MDSQs as

$$\delta_j = \frac{\delta}{a^j}. \tag{5.8}$$

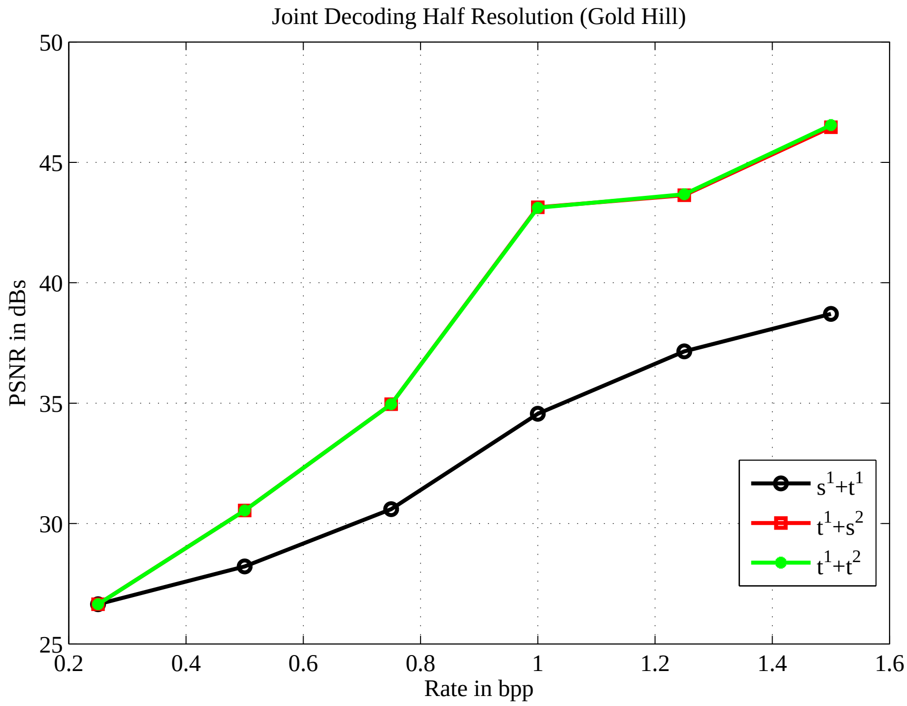
Figure 5.3 shows an example of the central and side quantizers for different resolution descriptions with parameters ($a = 2$, $f = 2$, $g = 2$, $r = 3$ and $P = 2$). As mentioned in Section 3.3 that not all the combinations from different MDSQs give the same quality improvement. Similar thing can be observed from Figure 5.3 and Table 5.1 that some combinations from different MDSQs give smaller implicit central quantizer bin width that results in better quality improvement, while the other combinations just contribute to improve the robustness of the overall system. Figure 5.4 shows the combinations that can be selected for different resolution descriptions in order to provide better rate distortion performance. The fully SMDC scheme for images is evaluated in two steps: Firstly, considering the transmission over lossless channel in order to study the rate distortion performance and secondly considering transmission along a packet erasure channel in order to evaluate the robustness of the scheme.

Table 5.2: Full resolution joint decoding quality (PSNR in dB) from different spatial resolution descriptions.

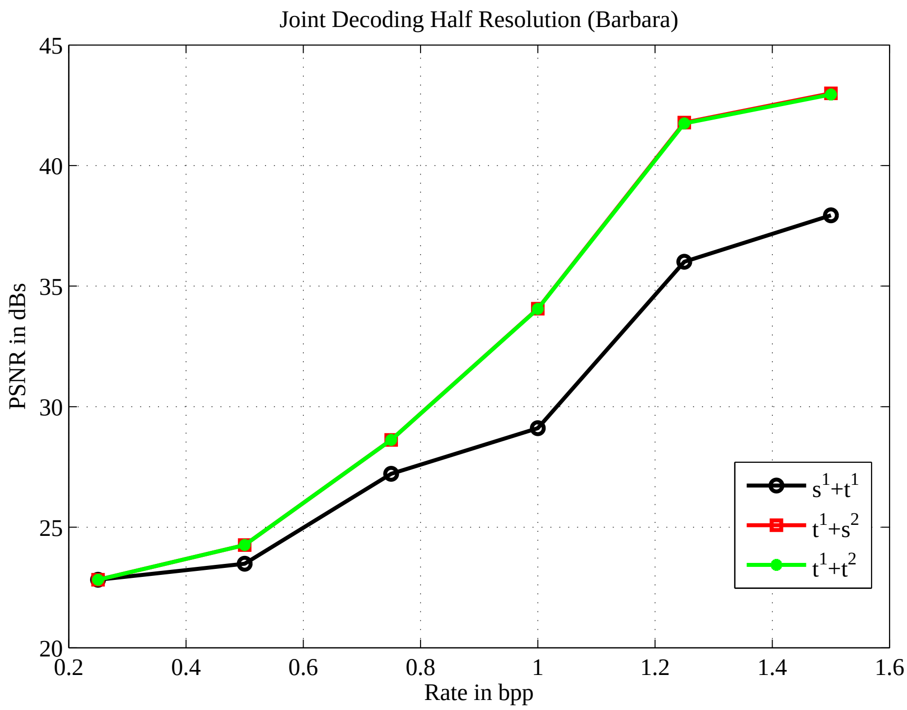
Test		D_{s^0,t^0}			D_{t^0,s^1}			D_{t^0,s^2}			D_{t^0,t^1,s^2}		
image #	a	$p=0$	$p=1$	$p=2$	$p=0$	$p=1$	$p=2$	$p=0$	$p=1$	$p=2$	$p=0$	$p=1$	$p=2$
1	2	27.13	34.05	40.54	27.36	34.32	40.78	27.81	35.13	41.37	28.40	35.15	41.33
1	3	27.13	34.05	40.54	27.50	34.42	40.91	28.36	35.21	41.42	28.66	35.28	41.48
2	2	27.15	34.30	40.76	27.78	34.96	41.23	28.41	35.82	41.68	28.54	35.33	41.48
2	3	27.15	34.30	40.76	28.02	35.14	41.28	29.45	35.93	41.77	29.24	36.02	41.83
3	2	29.70	36.28	42.50	30.98	37.01	43.00	31.54	37.46	43.28	32.11	37.84	43.49
3	3	29.70	36.28	42.50	31.29	37.18	43.18	32.57	37.78	43.45	32.65	38.03	43.73
4	2	31.34	37.46	42.17	33.04	38.44	42.68	33.28	38.61	42.81	34.35	39.22	43.05
4	3	31.34	37.46	42.17	33.82	38.76	42.78	34.41	38.94	43.15	34.95	39.37	43.25

The scalable descriptions generated by each MDSQ-SR have different spatial resolution *i.e.,* quarter half and full. There are different possibilities of sending and receiving different spatial resolution descriptions. The resolution of the decoded image is the same as the highest resolution among all the received descriptions. In the first set of simulation results, the distortion constraints defined in Eq. (5.1)-Eq. (5.3) are verified for different parameters of the MDSQ-SR for MUDC. Three MDSQ-SRs are considered to demonstrate the effect of joint decoding from different spatial resolution description for parameters $g_1 = g_2 = g_3 = 2$, $r = 3$ and $a = 2, 3$. Table 5.2 shows the joint decoding PSNR of different spatial resolution description, jointly decoded at full resolution for different MDSQ-SR parameters *i.e.,* $g_1 = g_2 = g_3 = 2$ for $a = 2, 3$ for four test images (#1 *Barbara (on chair)*, #2 *Barbara (on floor)*, #3 *Gold Hill*, #4 *Blackboard*). It is observed from Table 5.2 that the joint decoding quality of different spatial resolution description is better than the joint decoding quality of the same resolution description for all refinement levels p. The joint decoding quality improvement is only possible if the design conditions of the MUDC and MDSQ-SR proposed in Chapter 3 and Chapter 4 are satisfied *i.e.,* ($r > g$, r and g are not integer multiple of each other, and $a \geq g$). The joint decoding quality improvement of different spatial resolution description is due to different bin spread of the side quantizers of the base layer and different number of refinement levels of the enhancement layers.

In rate distortion performance, the joint decoding of two similar and different spatial resolution descriptions is considered. Each spatial resolution description created from MDSQ-SR is extracted according to the procedure mentioned in Section 5.1.1. Figure 5.5 and Figure 5.6 shows the rate distortion performance of the joint decoding of different resolution descriptions generated from different MDSQ-SRs at half and full spatial resolution respectively. It can easily be observed from Figure 5.5 and Figure 5.7 that the rate distortion performance improves if the descriptions from different resolution MDSQ-SR are jointly decoded. The best performance for joint decoding at full resolution is the one, in which full and quarter resolution description is jointly decoded. Similarly, the joint decoding of full and half spatial resolution descriptions gives better rate distortion performance than that of the joint decoding of full resolution descriptions. However, the rate distortion performance of the joint decoding of full and half resolution descriptions is less than the joint decoding of full and quarter resolution descriptions. The PSNR improvement of the joint decoding of the de-

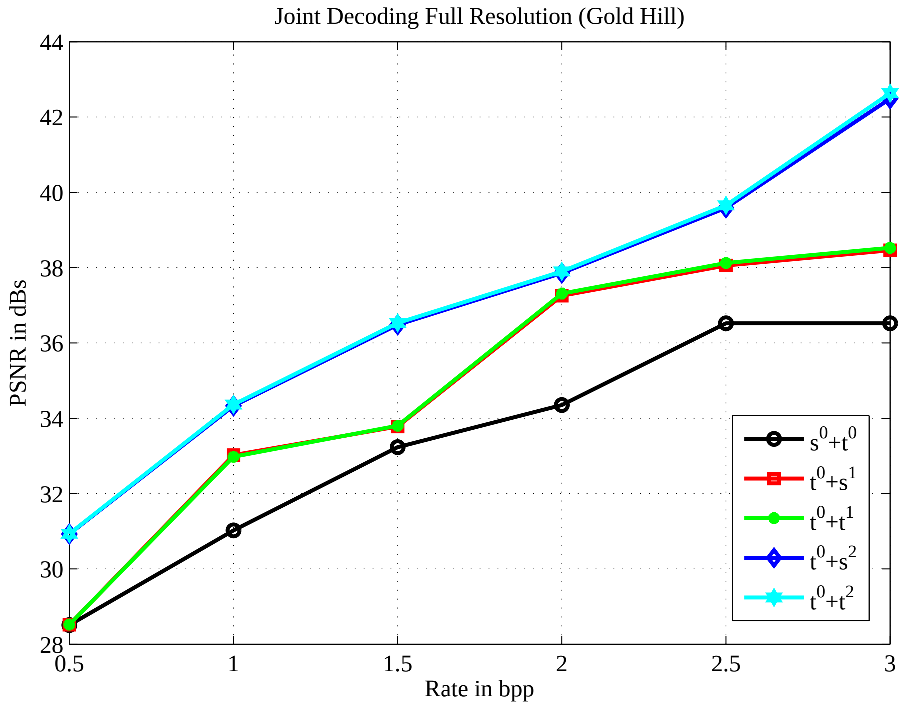


(a)

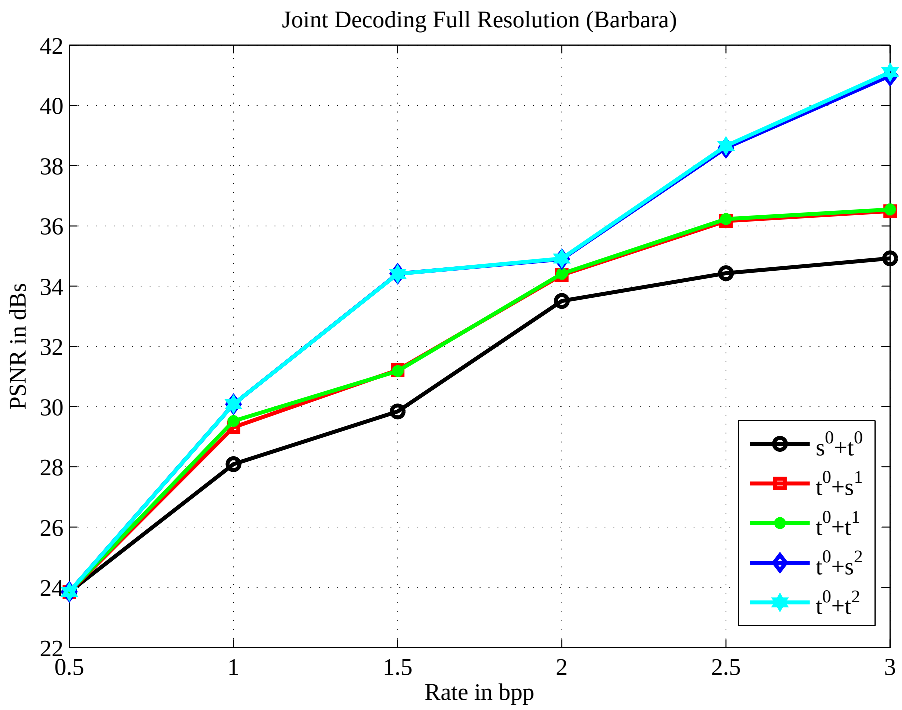


(b)

Figure 5.5: Rate distortion performance of joint decoding at half spatial resolution from different MDSQ-SRs for (a) *Gold Hill* (b) *Barbara (on chair)* images.



(a)



(b)

Figure 5.6: Rate distortion performance of joint decoding at full spatial resolution from different MDSQ-SRs for (a) *Gold Hill* (b) *Barbara (on chair)* images.

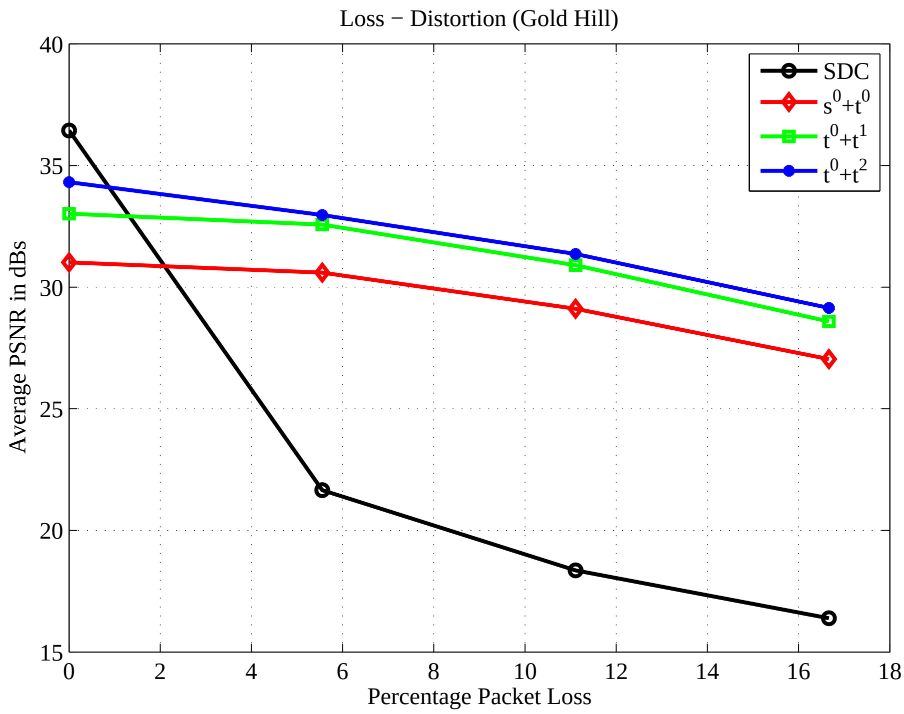
$s^0 + t^0$ 31.02 dB | $t^0 + s^1$ 33.02 dB | $t^0 + s^2$ 34.36 dB

$s^0 + t^0$ 28.08 dB | $t^0 + s^1$ 29.37 dB | $t^0 + s^2$ 30.08 dB

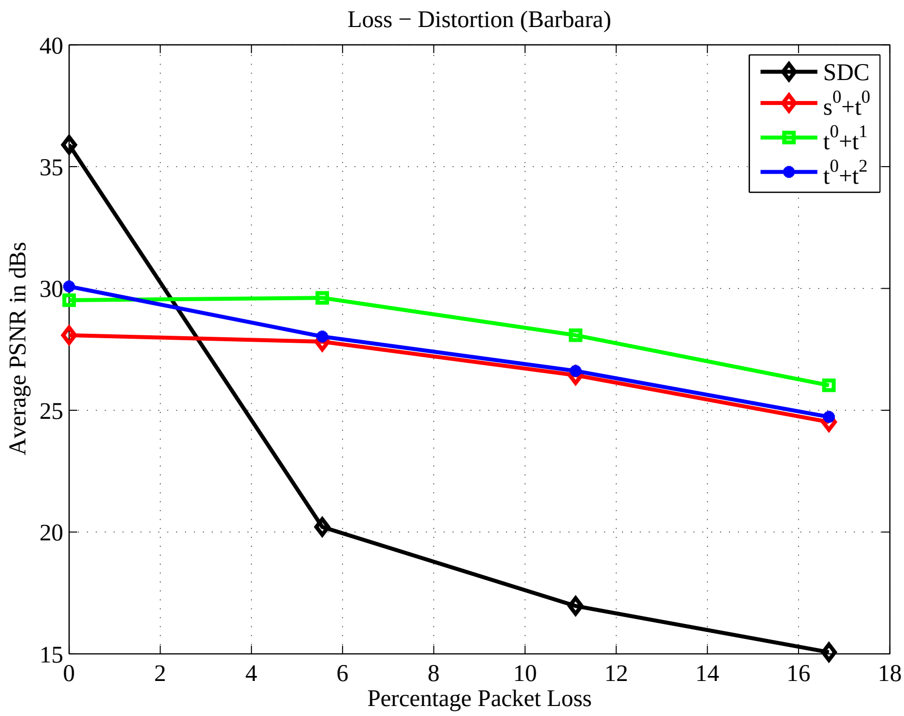
Figure 5.7: A part of the decoded *Gold Hill* (Row 1) and *Barbara (on chair)* (Row 2) images after joint decoding of different resolution description.

scriptions from different resolution MDSQ-SR is due to the availability of more number of refinement levels of the same subband at the decoder at same data rate than the number of refinement levels of the joint decoding from the same resolution MDSQ-SR.

Figure 5.7 shows the portion of the decoded *Gold Hill* (Row 1 : without packet loss, Row 2 : with 17% packet loss) and *Barbara (on chair)* (Row 3 : without packet loss, Row 4 : with 17% packet loss) images and their PSNR value after the joint decoding of different spatial resolution description, where each description is extracted at 0.5 bpp. It is evident from Figure 5.7 how the quality improves by joint decoding from different resolution description at the same data rate. The joint decoding from quarter and full resolution descriptions shows the PSNR improvement of 3.34 dB for *Gold Hill* and 2.0 dB for *Barbara (on chair)* image with respect to the joint decoding from full resolution descriptions. Similarly, the joint decoding from half and full resolution descriptions show the PSNR improvement of 2.0 dB for *Gold Hill* and 1.29 dB for *Barbara (on chair)* images



(a)



(b)

Figure 5.8: Effect of packet loss on PSNR for different percentage of packet drops for joint decoding at full spatial resolution for (a) *Gold Hill* (b) *Barbara (on chair)* images.

$s^0 + t^0$ 31.02 dB $\quad$ $t^0 + s^1$ 33.02 dB $\quad$ $t^0 + s^2$ 34.36 dB

$s^0 + t^0$ 28.13 dB $\quad$ $t^0 + s^1$ 29.35 dB $\quad$ $t^0 + s^2$ 28.95 dB

$s^0 + t^0$ 28.08 dB $\quad$ $t^0 + s^1$ 29.37 dB $\quad$ $t^0 + s^2$ 30.08 dB

$s^0 + t^0$ 26.15 dB $\quad$ $t^0 + s^1$ 26.79 dB $\quad$ $t^0 + s^2$ 25.76 dB

Figure 5.9: A part of the decoded *Gold Hill* and *Barbara (on chair)* after joint decoding of different resolution description without and with packet loss.

with respect to the joint decoding from full resolution descriptions.

For the last set of experiments, wavelet-based packetization is used for each description as mentioned in Section 3.4.2. For this set of experiments packets for each description is transmitted through a channel model having packet loss rates varying between 0% and 18%. Results of different combination for $N = 2$ descriptions from quarter, half and full resolution descriptions and single description coding under packet erasure channels are presented in Figure 5.8 for *Gold Hill* and Barbara images. It is easily observed that the decay of the PSNR is more for the joint decoding of descriptions from different spatial resolution MDSQ-SR than the decay of the joint decoding PSNR of descriptions from the same resolution MDSQ-SR. The increase in decay is due to the unavailability of coefficients from higher frequency subbands at the decoder when different resolution descriptions are jointly decoded.

Figure 5.9 shows the portion of the decoded *Gold Hill* and *Barbara (on chair)* images and their PSNR value after the joint decoding of different spatial resolution description without any packet loss and with 17% packet loss in each description, where each description is extracted at 0.5bpp. It is evident from Figure 5.9 that the decay in quality from the joint decoding from the same resolution descriptions is smaller than the joint decoding from different resolution descriptions. The PSNR decrement from joint decoding of both the full spatial resolution, half and full resolution and quarter and full resolution are 2.89 dB, 3.67 dB and 5.41 dB for *Gold Hill* and 1.93 dB, 2.58 dB and 4.32 dB for *Barbara (on chair)* images respectively. The decrement in PSNR is due to the missing subbands in quarter and half resolution descriptions.

5.2 Robustness Improvement of Base Layer

In this section a new method of making more robust base layer of scalable descriptions is presented that uses the concept of side information as used in distributed coding. In distributed coding, the correlation information present in the source is exploited at the decoder by using the side information. The side information can be extracted at the decoder by using guess, hint, learn or trial

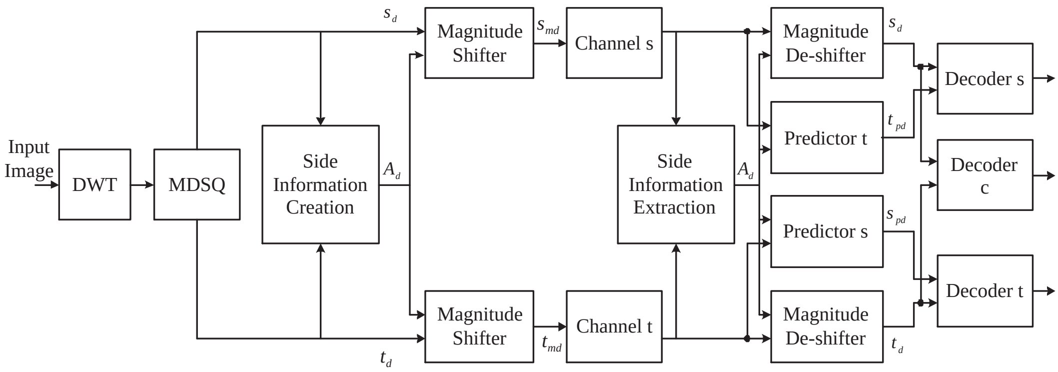


Figure 5.10: Block diagram of the multiple description image coding using side information.

approaches. On the other hand in MDC, correlated descriptions are generated to achieve better quality for joint decoding. The descriptions generated from the MDSQ have correlation information depending on the number of diagonals filled in the index assignment matrix and can be exploited at the decoder to improve the performance. Therefore, instead of transmitting each description generated by the MDSQ a magnitude shifted version of these descriptions are created by exploiting the correlation information, which in this case is the side information. The side information is embedded in the magnitude shifted version of each description. By using the side information to generate magnitude shifted version of the descriptions has two advantages.

1. It gives better side decoding quality, which is comparable to the central decoding quality of simple MDC.
2. Prediction of the data using side information gives high resilience under erroneous conditions making such a scheme more feasible for channels having high loss rate.

Figure 5.10 shows a block diagram of the multiple description image coding system using side information. In the proposed method, DWT is used to decompose the input image into different subbands. Each subband is quantized using different MDSQs depending on the importance of the information. Two descriptions are then generated based on the parameters of the MDSQ used for each subband. The generated descriptions from the MDSQ have some correlation information because the descriptions are leading or lagging from each other depending on the number of diagonals filled in the index assignment matrix. By using that lead and lag information, one description can be predicted from the other description. A magnitude shifted version of the description is obtained from the side quantizer of the MDSQ by using the side information that exploits the correlation information between the descriptions. The side information used in the proposed scheme is the absolute difference of the descriptions generated from the MDSQ. The details of the encoding and decoding procedure for such a system is given in next two subsections.

5.2.1 Encoding Procedure

Let s_d and t_d be the two descriptions generated from MDSQ d with central quatizer having n_d and side quantizers having m_d number of bins, depending on the factor f_d, where f is the number of diagonals filled in the index assignment matrix, d is the total number of subbands *i.e.,* $0 < d < 3G$, and G is the wavelet decomposition levels. The leading or lagging factor of the two descriptions generated from any MDSQ is 0 to $f_d - 1$. So the leading or lagging information can be exploited at the decoder to predict the missing description from the received one. In this method the side information considered is the absolute difference of the description generated from the MDSQ, which is written mathematically as,

$$A_d = \mid t_d - s_d \mid, \tag{5.9}$$

where $A_d \in \{0, 1, ..., f_d - 1\}$.

Let s_{sd} and t_{sd} be the magnitude shifted version of the description s_d and t_d. Once the side information is generated, the magnitude shifted version of the description based on the side information is obtained by using Eq. (5.10) and Eq. (5.11).

$$s_{sd} = s_d + \alpha_d A_d, \tag{5.10}$$

$$t_{sd} = t_d + \alpha_d A_d, \tag{5.11}$$

where α_d is the magnitude shifting factor and the value of α_d is greater than the maximum number of rows or columns in the index assignment matrix d *i.e.,*

($\alpha_d > m_d$). The magnitude shifted version of each description is entropy coded and transmitted through different communication channels.

5.2.2 Decoding Procedure

At the decoder, the side information is extracted from any of the received description by using Eq. (5.12) and Eq. (5.13) as

$$A_d = s_{md} - \beta\alpha_d, \tag{5.12}$$

$$A_d = t_{md} - \beta\alpha_d, \tag{5.13}$$

where value of β is selected in such a way that $A_d \in \{0, 1, ..., f_d - 1\}$. Once the side information matrix A_d is extracted either from s_{md} or t_{md}, the magnitude deshifting is performed to get s_d or t_d by using Eq. (5.14) and Eq. (5.15) respectively as,

$$s_d = s_{md} - \alpha_d A_d, \tag{5.14}$$

$$t_d = t_{md} - \alpha_d A_d, \tag{5.15}$$

If some part of a description is corrupted or dropped during transmission then the missing information is predicted from the received information, consequently increasing the overall robustness of the system. On side decoding, when single description is received at the decoder, the missing description is predicted from the magnitude deshifted version of the received description and the side information matrix. Let s_{pd} and t_{pd} be the predicted missing description and are obtained by using Eq. (5.16) and Eq. (5.17) respectively.

$$s_{pd} = t_d - A_d, \tag{5.16}$$

$$t_{pd} = s_d + A_d. \tag{5.17}$$

On single description decoding, the proposed method uses the magnitude deshifted

Table 5.3: Rate and PSNR differences of the side decoding of the base layer of MDSQ with and without side information.

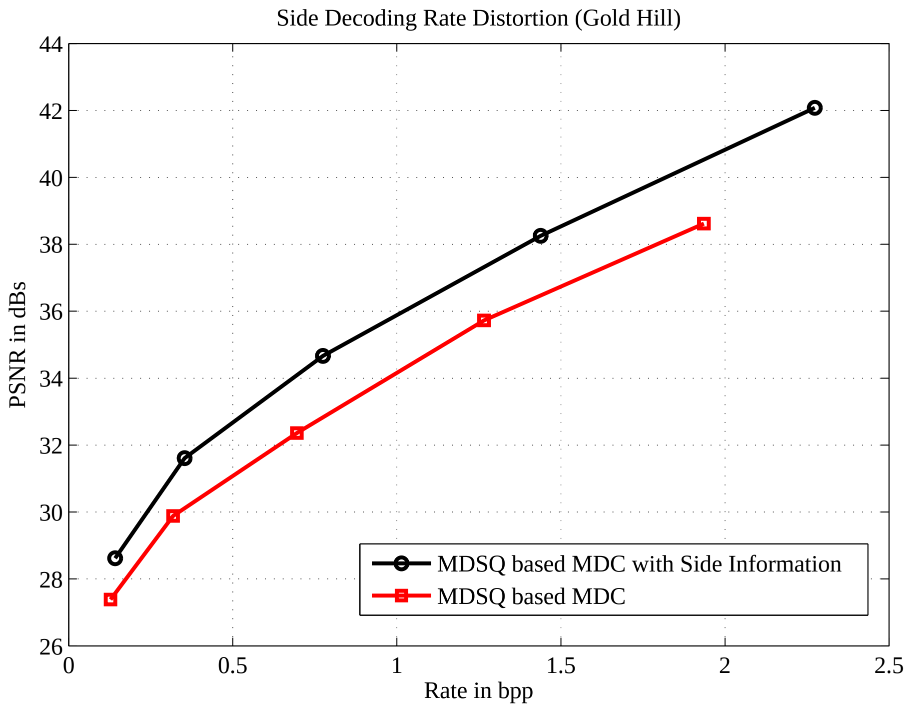
Test image #	Side Information Subbands $(0-9)$		Side Information Subbands $(0-12)$		Side Information Subbands $(0-15)$	
	Rate in bpp	PSNR in dBs	Rate in bpp	PSNR in dBs	Rate in bpp	PSNR in dBs
1	0.02	0.42	0.03	1.12	0.04	1.93
2	0.03	0.30	0.07	1.10	0.10	3.51
3	0.02	0.32	0.05	0.94	0.07	2.51
4	0.01	0.50	0.02	0.97	0.03	1.79

and prediction description in a similar way as used in central decoding. In joint decoding, only magnitude deshifting is used before the central decoder. The redundancy increase due to side information is controlled by generating the side information matrix for fewer subbands.

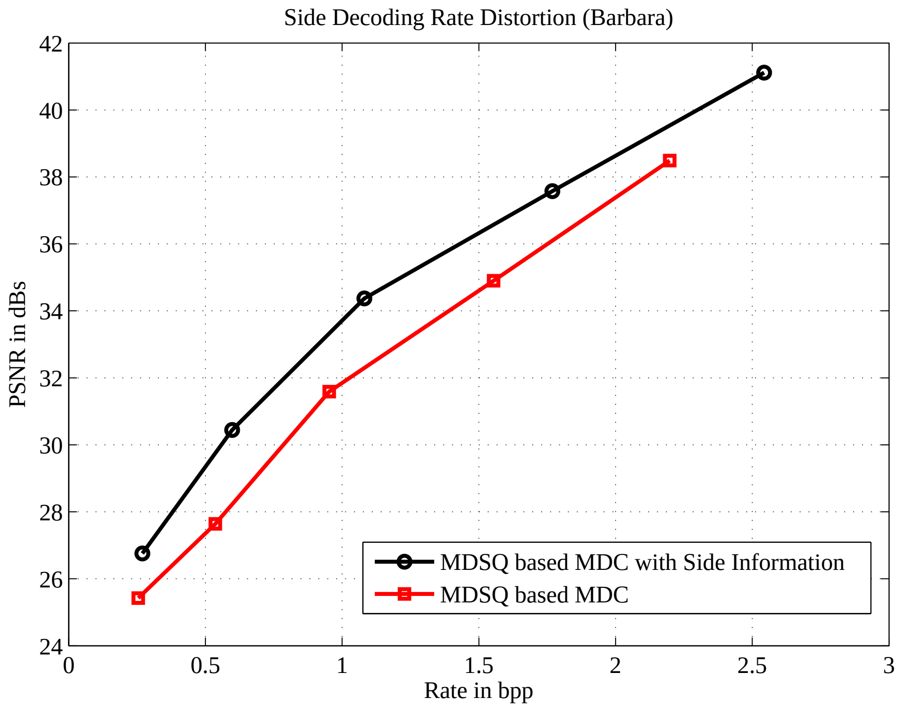
5.2.3 Performance Evaluation

The main aim of using side information in MDC framework is to improve the side description decoding and increase the overall robustness under lossy channel conditions. The proposed MDC scheme based on MDSQ with side information is evaluated in two steps: Firstly, considering transmission over a lossless channel in order to study the rate distortion performance and secondly considering transmission along a packet erasure channel in order to evaluate the robustness of the scheme.

Table 5.3 shows the difference in the rate and PSNR of the side description decoding from MDSQ-based MDC without and with side information for different subbands. It is clear from the Table 5.3 that by increasing the number of subbands for the side information, the side decoding PSNR increases at the cost of more data rate. Figure 5.11 and Figure 5.12 shows the comparison of the side and joint decoding rate distortion performance of the MDSQ based MDC scheme using side information and simple MDSQ-based MDC scheme for *Gold Hill* and Barbara images. It can be seen that the side description PSNR of the MDC scheme with side information is better than that of the simple MDSQ-based MDC scheme.

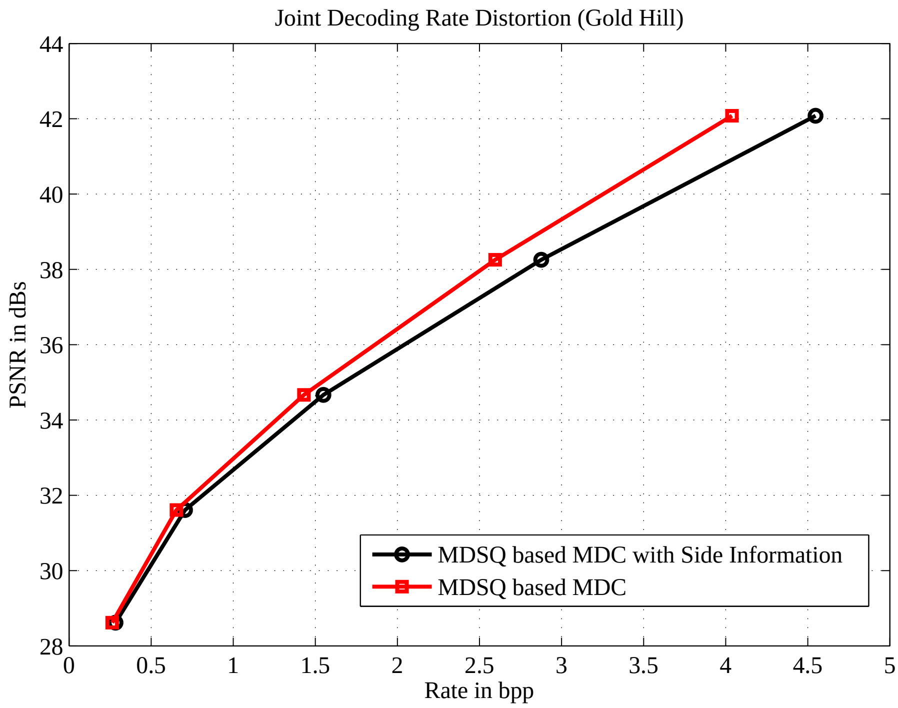


(a)

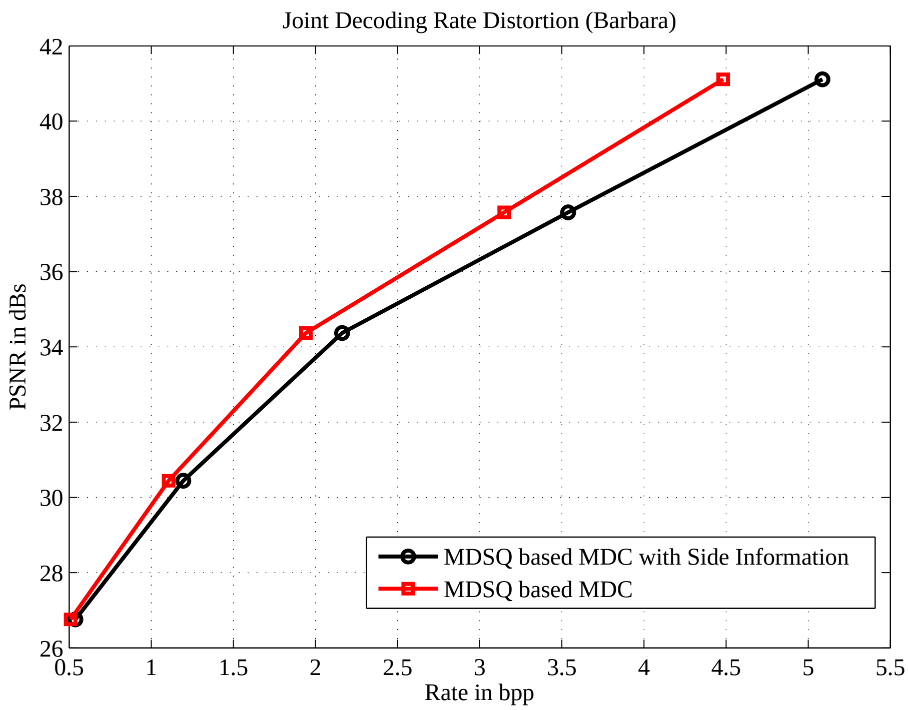


(b)

Figure 5.11: Side decoding rate distortion comparison of the MDSQ based MDC with and without using side information for (a) *Gold Hill* (b) *Barbara (on chair)* images.

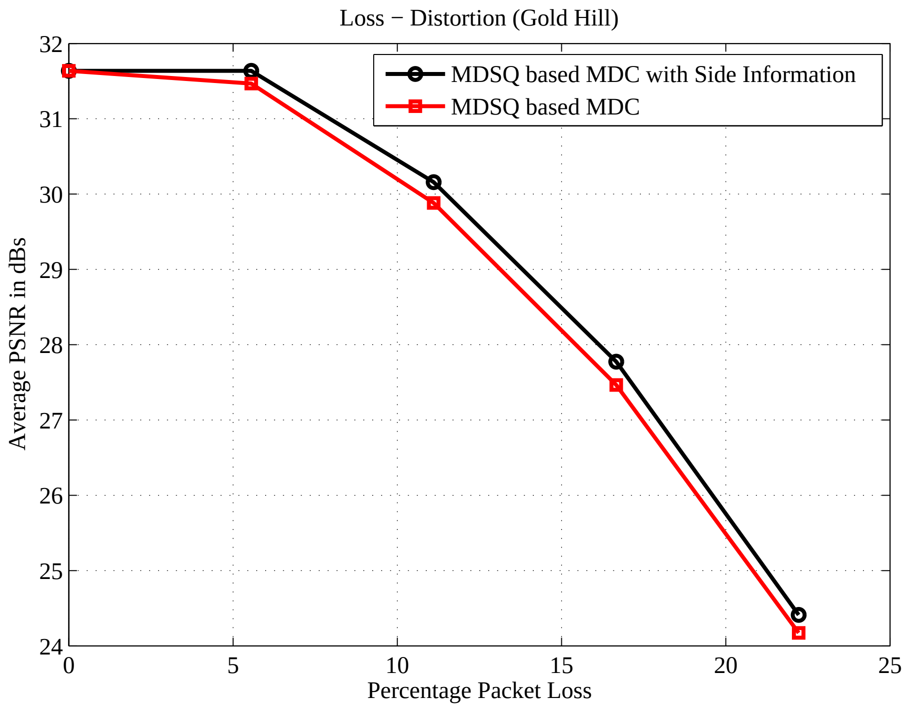


(a)

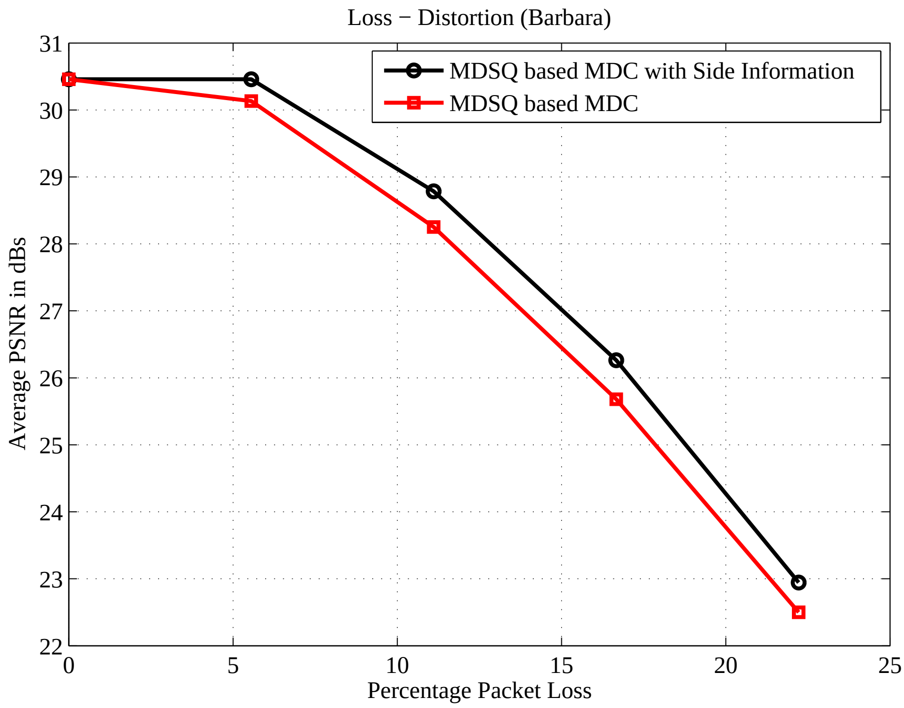


(b)

Figure 5.12: Joint decoding rate distortion comparison of the MDSQ based MDC with and without using side information for (a) *Gold Hill* (b) *Barbara (on chair)* images.



(a)



(b)

Figure 5.13: Performance evaluation comparison of the MDSQ based MDC using side information and the simple MDSQ based MDC over packet erasure channel for (a) *Gold Hill* (b) *Barbara (on chair)* images.

(a) (b)

Figure 5.14: Portion of the decoded Barbara image encoded at 0.5 bpp/channel after 22% packet loss in each description, (a) MDSQ based MDC using side information (b) Simple MDSQ based MDC.

Furthermore, the decoding quality of the side description is almost the same as the joint decoding quality of the simple MDSQ-based MDC scheme. On the other hand, the rate distortion curve of the joint decoding for the MDC using side information is below the curve of the simple MDC, representing the cost of the side information. The increase in the data rate is due to the side information that makes the MDC system more reliable under erroneous conditions, which is shown in second set of simulations.

The goal of the second set of simulation is to evaluate the performance of the proposed scheme over a packet erasure channel. For this purpose a similar kind of wavelet-based packetization is used as described in Section 3.4.2. Figure 5.13 shows the comparison of the MDC scheme using side information and simple MDC scheme based on MDSQ under packet erasure channel. The decoding quality is improved by 0.2-0.7 dBs under lossy channel conditions by using the side information in MDSQ-based MDC framework. The PSNR improvement is due to the prediction of the missing description from the received one. Figure 5.14 shows a portion of the decoded Barbara image after encoding multiple descriptions with and without side information in MDSQ-based MDC when 22% of packets are lost in each description. The superior visual quality of MDSQ-based MDC scheme with side information of the decoded Barbara image is evident in Figure 5.14 when compared with simple MDSQ-based MDC scheme.

5.3 Summary

In this chapter, two different frameworks for multiple description image coding using MDSQ were presented. The objective of the first framework was to achieve both the quality and resolution scalability in multiple description image coding by integrating the MUDC and MDSQ-SR schemes. Different spatial resolution descriptions were generated from several MDSQ-SRs in such a way that the distortion constraints defined in Chapter 3 were satisfied. Similarly, each MDSQ-SR for different spatial resolution description was designed according to the distortion constraints mentioned in Chapter 4. The joint decoding of the description from different resolution MDSQ-SR has shown better rate distortion performance than the joint decoding of the descriptions from the same MDSQ-SR, if the design conditions of MUDC and MDSQ-SR are satisfied. It is observed that the joint decoding from quarter and full resolution descriptions shows the PSNR improvement of 2.65 dB with respect to the joint decoding from full resolution descriptions. Similarly, the joint decoding from half and full resolution descriptions shows the PSNR improvement of 1.6 dB with respect to the joint decoding from full resolution descriptions. On the other hand, the PSNR of the joint decoding of different resolution description has shown a rapid decay than the PSNR of joint decoding of similar resolution description under packet erasure channel. The decay in the PSNR value is due to the unavailability of higher frequency subbands in lower resolution description. However, the joint decoding PSNR from any spatial resolution description has shown much better performance than the single scalable image coding.

The main aim of the second framework proposed in this chapter was to increase the side decoding quality of the base layer and overall robustness of the fully SMDC system by using the concept of side information in multiple description image coding. The side information, which we used, is the absolute difference of the descriptions generated from the MDSQ. Simulation results of the side information-based MDC has shown almost the same side decoding quality as central decoding. On the other hand, by using the side information in MDSQ-based base layer the performance of the joint decoding is improved by 0.2-0.7 dB under packet loss environment.

Chapter 6

Scalable Multiple Description Video Coding

So far, in this thesis methods for generating any number of scalable descriptions for images are presented. This chapter focuses on scalable multiple description video coding (SMDVC) to achieve quality, resolution and frame rate scalabilities. The temporal or frame rate scalability is achieved by using MCTF, while the quality and resolution scalability is achieved in a similar way as achieved in MDC for images by using MDSQ-SR and MUDC respectively. The rest of the chapter is organized as follows: The proposed scheme for SMDVC is presented in Section 6.1. The redundancy control mechanism for texture and motion vector information is presented in Section 6.2. Section 6.3 evaluates the proposed SMDVC scheme for P2P streaming followed by the summary in Section 6.4.

6.1 MDSQ-SR-based Multiple Description Video Coding

In Section 2.2 it is mentioned that any scalable content coding framework is divided into three blocks *i.e.,* encoder, extractor and decoder. Therefore, any SMDVC solution consists of these three blocks. In SMDVC framework, the en-

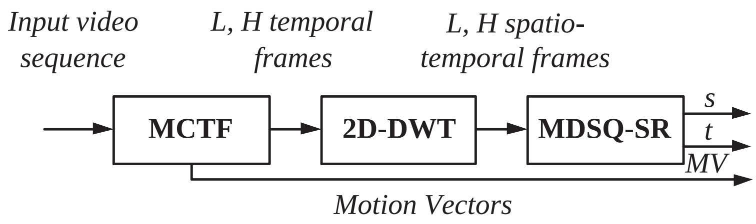


Figure 6.1: Block diagram of proposed SMDVC encoder.

coder is responsible for generating two or more scalable bitstreams called descriptions and information related to the structure of each description. The maximum reconstructional quality is achieved at the decoder if all the information from each scalable description is received. However, it is not always necessary to send all the information from each description, especially when the transmission over heterogeneous networks is considered. Therefore, extractor block is required to truncate the scalable descriptions according to the parameters such as quality, resolution and frame rate to extract the descriptions at required rate. Finally, the decoder is capable of decoding the extracted descriptions individually and jointly.

The SMDVC scheme presented in this chapter is based on MCTF and MDSQ-SR. The block diagram of the proposed SMDVC encoder is shown in Figure 6.1. Firstly, the MCTF is performed on input sequence to remove the temporal correlation among the frames. The input sequence frames are decomposed into low pass, high pass temporal frames and set of motion vectors by temporally splitting the input video into even and odd frames followed by motion compensation and filtering. The decomposed low pass temporal frames (L-frames) and high pass temporal frames (H-frames) are further decomposed by repeating the MCTF operation on L-frames until the desired number of temporal decomposition levels is achieved. After temporal decomposition, each L and H frame is spatially decomposed by 2D-DWT to remove the correlation within the frame. MCTF and 2D-DWT completes the spatio-temporal decomposition of the input sequence. The spatio-temporal decomposition after three levels of temporal and one level of spatial decomposition is shown in Figure 6.2.

Two scalable descriptions, s and t, that provides robustness to the transmission error are created by quantizing the spatio-temporal coefficients using MDSQ-SR for P refinement levels. The main aim of MDSQ-SR is to create two scalable

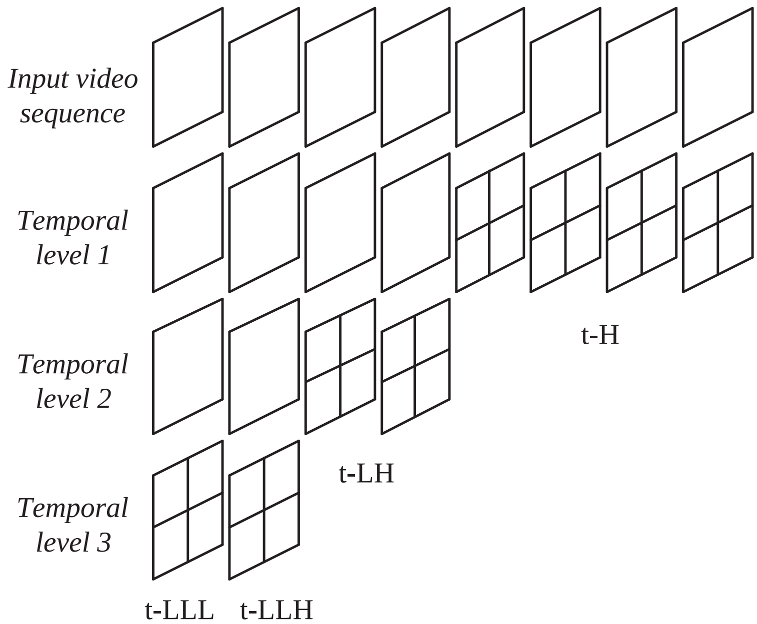


Figure 6.2: Frames after three temporal and one spatial level decomposition for the GOP size of 8 frames.

descriptions, which are capable of joint decoding using equal and unequal number of refinement levels from each of the side quantizers leading to balanced and unbalanced descriptions scenarios respectively. In MDSQ-SR, the base layer side quantizers are created by MDSQ, which are then successively refined to guarantee the distortion reductions leading to the enhancement layers of the side quantizers. The amount of redundancy between the descriptions can be controlled by using different index assignment strategies for the base layer.

In MDSQ-SR, each bin of the side quantizer is refined by splitting into r portions leading to refinement side quantizers, resulting in the quality enhancement side description layers. For MDSQ-SR the distortion of side quantizer refined up to any quantization level p, should be less than that for the side quantizer refined up to the quantization level $p-1$. Similarly, the distortion of the joint decoding at any quantization level p, should be less than the individual decoding of the side quantizer at level p and less than the joint decoding distortion refined up to the quantization level $p-1$. The distortion of descriptions created from MDSQ is related to each other depending on the factor g, *i.e.,* the maximum side quantizer bin spread depending on number of diagonals filled in the index assignment matrix of the base layer. In Chapter 4, it is proposed that $r > g$ and r and g are not integer multiple of each other to fulfil the joint and side qunaitzer refinement distortion constraints. For the proposed SMDVC scheme, it is assumed that motion vectors are repeated and embedded in each description and are transmitted securely, therefore available at the decoder without any losses.

6.1.1 Rate-Distortion-based Extraction for SMDVC

Once the scalable descriptions s and t are encoded by using MDSQ-SR, the extractor block extracts the descriptions according to the data rate requirement by selecting subset of temporal levels, spatial subbands and refinement levels of the MDSQ-SR according to the frame rate, resolution and quality preferences. The description s and t are extracted in such a way to minimize the joint distortion at required rate. For this purpose, the distortion expression is required for complete video sequence that considers all the contribution parameters, like temporal and spatial decomposition subbands and refinement levels of the MDSQ-SR. For any discrete signal and its approximation, the distortion is measured by the MSE and is calculated by using Eq. (2.1).

Let N be the total number of frames in a video sequence v, and E and G are the temporal and spatial decomposition levels respectively. Let V denote the spatio-temporal decomposition of the video sequence v. Similarly, $\hat{V}$ and $\hat{v}$ be the quantized version of spatio-temporal coefficients and the corresponding decoded video sequence after inverse spatial and temporal transform respectively. The quantization is the only contributing factor to the distortion associated to $\hat{v}$, as the inverse transformation are perfectly invertible. The distortion measured in spatial and frequency domain is the same if both the temporal and spatial transforms are orthonormal. For non orthonormal transforms the distortion in transform domain is an approximation of the distortion in spatial domain. For extraction purposes, the distortion in spatio-temporal domain is considered to avoid the inverse spatial and temporal transformation for each successive refinement. The distortion of video in transform domain by considering each frame of every GOP is written as,

$$D(\hat{V}) = \frac{1}{N} \sum_{i=0}^{N-1} D(\hat{V}_e). \tag{6.1}$$

Let B be the number of GOPs in video sequence and 2^E be the number of frames in each GOP. Then the distortion contribution considering each frame in each GOP is,

$$D(\hat{V}) = \frac{1}{2^E B}\sum_{b=0}^{B-1}\sum_{e=0}^{2^E-1} D(\hat{W}_{b,e}), \tag{6.2}$$

where $\hat{W}_{b,e}$ is the temporal decomposed frame e in GOP b. Let $\hat{W}_{b,e,d}$ be the subband after spatio-temporal decomposition, where $d = 0, \cdots, 3G$. Now Eq. (6.2) becomes,

$$D(\hat{V}) = \frac{1}{2^E B}\sum_{b=0}^{B-1}\sum_{e=0}^{2^E-1}\sum_{d=0}^{3G} D(\hat{W}_{b,e,d}). \tag{6.3}$$

Two scalable descriptions are created by using MDSQ-SR for P refinement levels, then the contribution of each spatio-temporal subband for each refinement level p to the total distortion is,

$$D(\hat{V}) = \frac{1}{2^E B}\sum_{b=0}^{B-1}\sum_{e=0}^{2^E-1}\sum_{d=0}^{3G}\sum_{p=0}^{P-1} D(\hat{W}_{b,e,d,p}), \tag{6.4}$$

The distortion considered in Eq. (6.4) is only by the quantization of the texture information in each temporally and spatially decomposed frame. There is no distortion term related to motion vectors because it is assumed that the motion vector information is available at the decoder without any loss. Similarly, the rate contribution for each temporal subband, spatial subband and successive refinement level to overall rate is,

$$R(V) = \sum_{b=0}^{B-1}\sum_{e=0}^{2^E-1}\sum_{d=0}^{3G}\sum_{p=0}^{P-1} R_{b,e,d,p}. \tag{6.5}$$

Let R_s and R_t be the bit rate requirement of each extracted scalable description, then the description s and t are extracted in such a way to minimize the joint distortion D_c at required rate *i.e.,* $R_c = R_s + R_t$ by selecting different spatio-temporal subbands and refinement levels. This can be written mathematically as in Eq. (6.6).

$$min\ D_c(\hat{V})\ subject\ to\ R_c \leq R_s + R_t. \tag{6.6}$$

The distortion expression in Eq. (6.4) considers all the temporal frames and spatial subbands, but different refinement levels p in each description s and t.

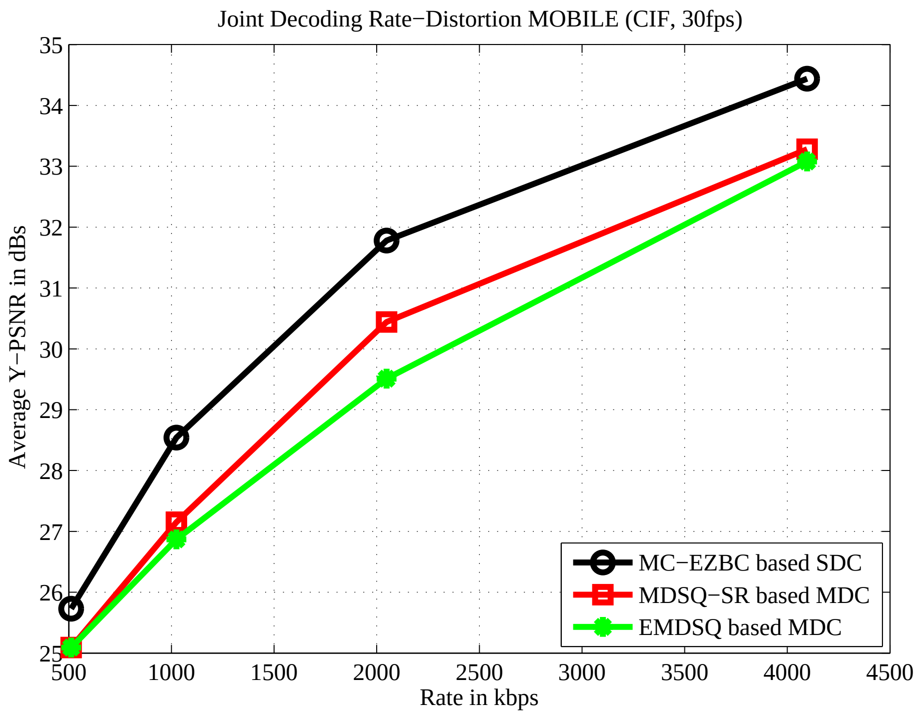
The rate distortion performance of the proposed SMDVC can be improved by considering different temporal frames for each description s and t. This type of scheme is discussed in Section 6.2.

6.1.2 Simulation Parameters and Results

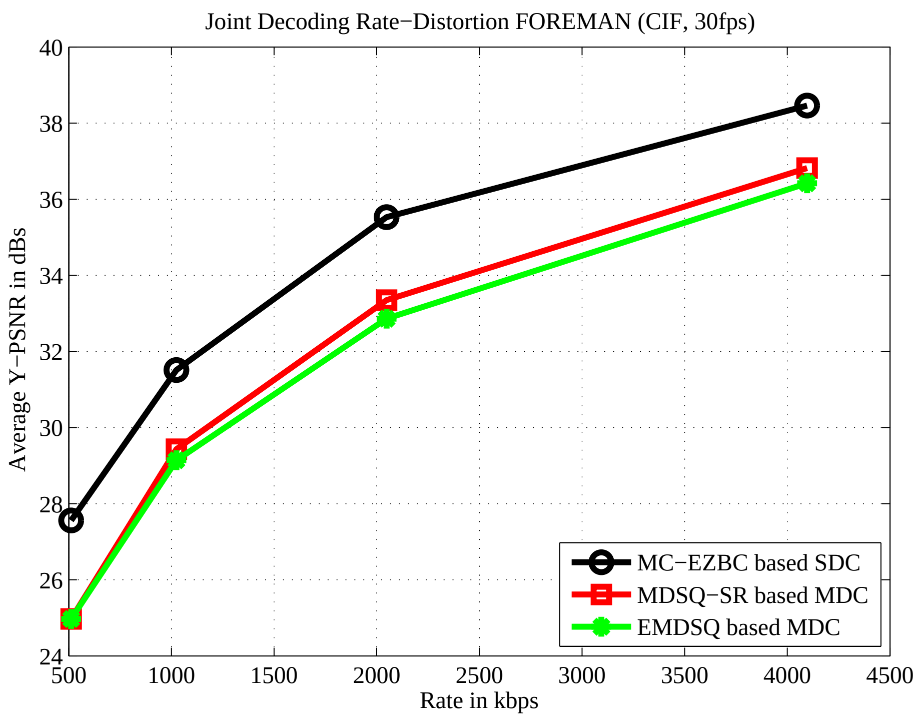
For the simulations, $N = 2$ descriptions is considered, where each scalable description is generated from MDSQ-SR after four levels of temporal *i.e.,* $E = 4$ and five levels of spatial *i.e.,* $G = 5$ decompositions. The number of successive refinement levels for each scalable description generated by MDSQ-SR is set to $P = 3$. The parameters related to each MDSQ-SR are $f = 2$ and $r = 3$, where f is the number of diagonals filled in the index assignment matrix and r is the successive refinement factor of each side quantizer bin. The simulation results are presented in two stages: performance under lossless and lossy channel conditions.

6.1.2.1 Performance under Lossless Channel Conditions

The rate distortion performance of the joint and individual decoding of the proposed MCTF and MDSQ-SR-based SMDVC at full frame rate is shown in Figure 6.3 and Figure 6.4 respectively. The joint decoding of the MDSQ-SR-based SMDVC is also compared with the MC-EZBC-based single scalable description video coding (SSDVC) and EMDSQ-based MDVC. The joint decoding PSNR of the proposed scheme is less than the SDDVC because of the redundancy introduced within the two scalable descriptions. However, the joint decoding performance of the MDSQ-SR-based SMDVC is better than the joint decoding performance of the EMDSQ-based MDVC scheme. At low data rates both the MDSQ-SR and EMDSQ-based MDVC schemes show the same joint decoding performance due to the same base layer and motion vector repetition. However, the MDSQ-SR based SMDVC gives on average of 0.32 dB better joint decoding performance than the joint decoding performance of the EMDSQ-based scheme at the same level of redundancy. The redundancy between the descriptions can be decreased by reducing the frame rate of individual descriptions, which is considered in next section.

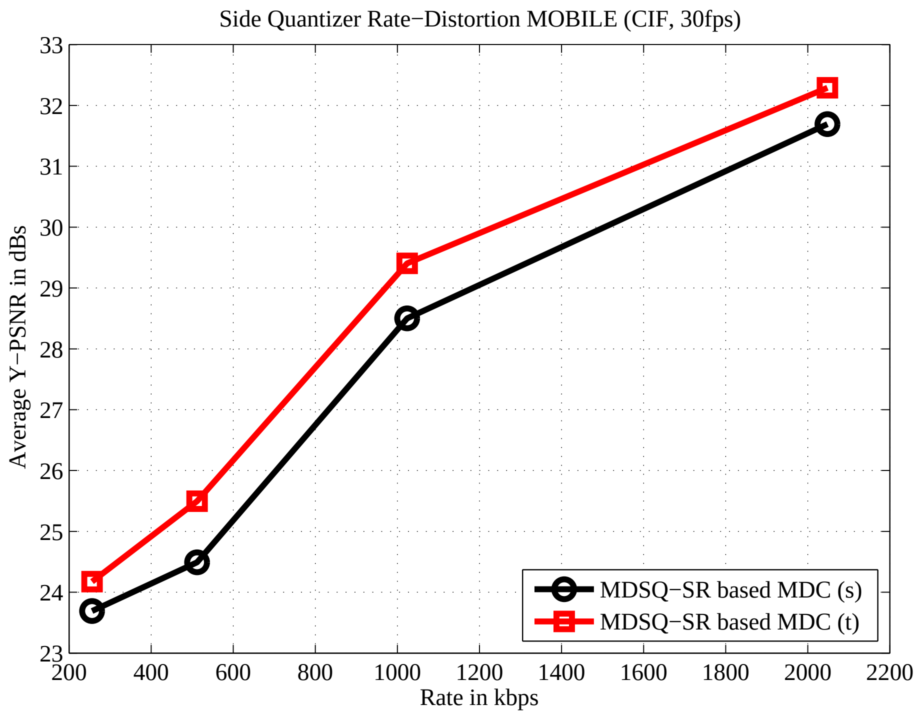


(a)

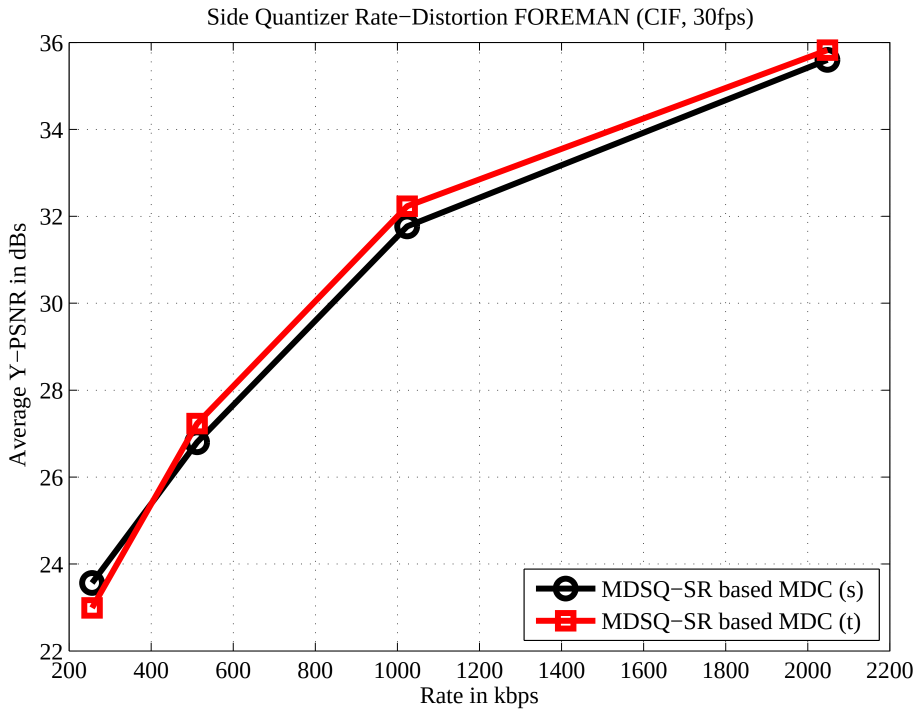


(b)

Figure 6.3: Rate distortion comparison of joint decoding of MDSQ-SR based SMDVC and MC-EZBC based SDC for (a) *Mobile* (b) *Foreman* sequences.

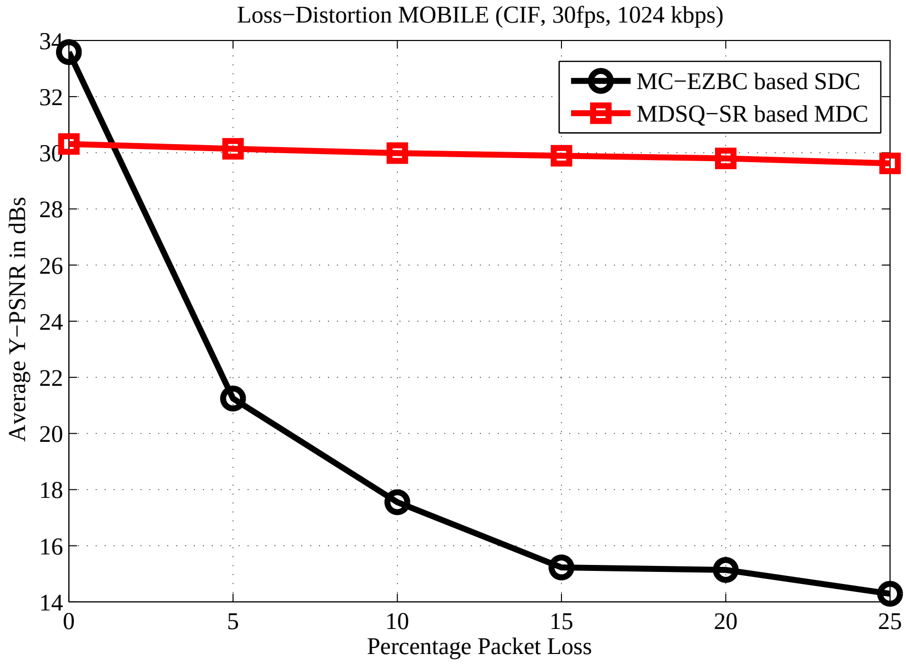


(a)

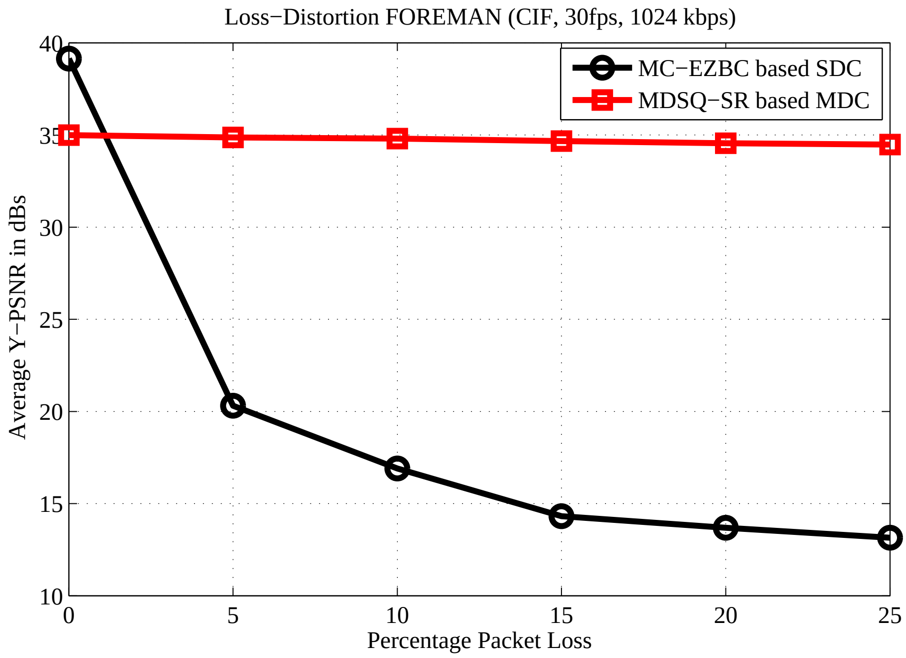


(b)

Figure 6.4: Rate distortion comparison of individual decoding of MDSQ-SR based SMDVC and MC-EZBC based SDC for (a) *Mobile* (b) *Foreman* sequences.



(a)



(b)

Figure 6.5: Loss-Distortion performance due to packet loss for (a) *Mobile* and (b) *Foreman* sequence encoded at 1024 kbps.

Figure 6.6: Portion of the decoded frame 12 of *Foreman* sequence after 25% packet loss from MC-EZBC SDVC (left) and MDSQ-SR based SMDVC (right).

6.1.2.2 Performance under Lossy Channel Conditions

For the second set of simulations, wavelet tree-based packetization is used. The coefficients of each spatio-temporal subband and its refinement levels are placed in the same packet to minimize the erasure effect. The average PSNR for particular number of packet loss is calculated by averaging the PSNR for all iterations. Performance under packet erasure channel is evaluated by randomly choosing 100 different packet loss patterns for each description from the total packet loss patterns as generated in Section 3.4.2. Figure 6.5 shows the comparison of the MDSQ-SR-based SMDVC and MC-EZBC-based SDVC scheme for *Mobile* and *Foreman* sequences, encoded at 1024 kbps under packet loss environment. It is easily observed that the SDVC performance deteriorates rapidly as the percentage of packets loss increases. However, the proposed SMDVC shows high robustness against packet losses. The superior visual quality of the MDSQ-SR-based SMDVC with respect to that of the MC-EZBC-based SDVC for the 512 kbps data rate with 25% packet loss rate is demonstrated in Figure 6.6.

6.2 Redundancy Control in SMDVC

In Section 6.1, SMDVC based on MCTF and MDSQ-SR is presented that considers all the temporal and spatial subbands for each description, but different number of refinement levels of the side quantizers in each description for side and central decoding at some particular data rate. Also the motion vector information is repeated in each description. SMDVC scheme results in high redundancy or bit budget by repeating motion vector information and considering all temporal frames in each description. The amount of redundancy between the descriptions can be controlled by reducing the texture or motion vector information in each description.

The bit budget or cost of the scalable description s and t is represented R_s and R_t respectively and can be divided into two parts *i.e.,* texture and motion vector cost. Let $R_{s_{txt}}$, $R_{t_{txt}}$ and $R_{s_{MV}}$,$R_{t_{MV}}$ be the cost of texture and motion vector in description s and t respectively. The bit budget requirement of each description is reduced by either reducing the cost of the texture or motion vector information. In the proposed scheme, the texture cost is reduced by splitting high frequency temporal frames in each description while the motion vector cost is reduced by creating multiple motion vector description using MDSQ.

6.2.1 Temporal Frame Splitting

The performance of the joint and side decoding depends on the amount of redundancy between the descriptions. The joint decoding quality of the SMDVC scheme proposed in Section 6.1 at required bit budget can be enhanced by splitting the temporal subbands into two descriptions. In the proposed SMDVC, different number of refinement levels of the side quantizer p are selected for all temporal and spatial subbands in each description at required rate R_s and R_t. In temporal frame splitting based SMDVC scheme, the frames in the last temporal decomposition level are repeated in each description as the low temporal levels has more energy distribution. The rest of the frames from temporal levels are splitted into two descriptions by selecting even and odd frames for description s and t respectively. Figure 6.7 shows the GOP structure of spatio-temporal sub-

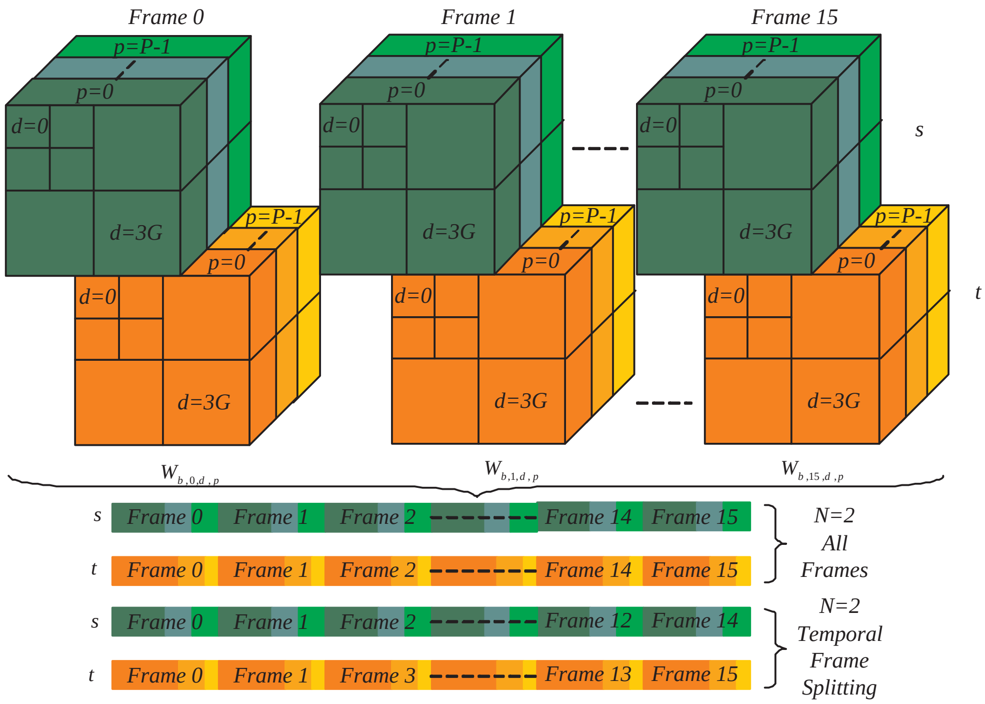


Figure 6.7: Selection criteria for description s and t for temporal frame splitting based scheme.

bands with P levels of refinement of MDSQ-SR and how the two descriptions are created by considering different frames in each description. The contribution of each spatio-temporal subband for each refinement level p to the total distortion for description s and t are

$$D_s(\hat{V}) = \frac{1}{2^E B} \sum_{b=0}^{B-1} \sum_{e_s} \sum_{d=0}^{3G} \sum_{p=0}^{P-1} D(\hat{W}_{b,e_s,d,p}), \tag{6.7}$$

$$D_t(\hat{V}) = \frac{1}{2^E B} \sum_{b=0}^{B-1} \sum_{e_t} \sum_{d=0}^{3G} \sum_{p=0}^{P-1} D(\hat{W}_{b,e_t,d,p}), \tag{6.8}$$

where $e_s \in \{0, 1, 2, 4..., 2^E - 1\}$ and $e_t \in \{0, 1, 3, 5..., 2^E - 1\}$. Similarly, the rate contribution for each temporal, spatial subband and refinement level to overall rate is,

$$R_c(V) = \sum_{b=0}^{B-1} \left(\sum_{e_s} \sum_{d=0}^{3G} \sum_{p=0}^{P-1} R_{b,e_s,d,p} + \sum_{e_t}^{2^E-1} \sum_{d=0}^{3G} \sum_{p=0}^{P-1} R_{b,e_t,d,p} \right). \tag{6.9}$$

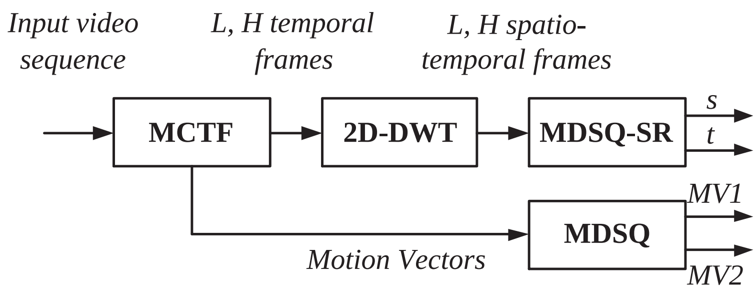


Figure 6.8: Block diagram of SMDVC using MDSQ-SR with multiple motion vector streams using MDSQ.

By using the frame splitting in temporal domain the number of refinement level p of the side quantizer is increased for each description at required rate R_s and R_t. At the decoder, on joint decoding the missing temporal frames in one description is interleaved from other descriptions before the inverse quantization and transform operation. On the other hand, on side decoding the missing frame is assumed as zero spatio-temporal coefficients. The motion vector generated from the MCTF block is repeated in each description as done in SMDVC presented in the previous section.

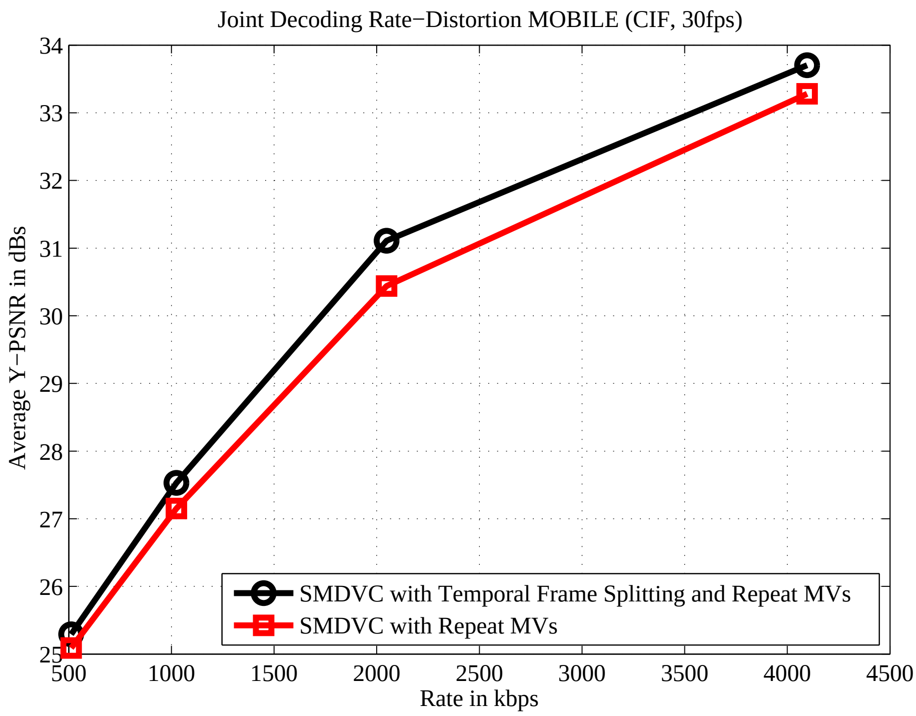
6.2.2 Multiple Motion Vectors using MDSQ

In most of the MDVC methods, the same motion vector information is used in each description to avoid the mismatch. By repeating the motion vector information in each description not only increases the rate and redundancy among the descriptions, but also there is no advantage in terms of distortion improvement under lossless channel conditions, when both the motion vector streams are available at the decoder. Therefore, for a complete MDVC system, both the texture and motion vector information are encoded in multiple fashion. As MDSQ-SR is used to create two scalable descriptions of the texture information, the MDSQ can be used to create multiple descriptions of the motion vector information. The motion vector streams are created in such a way to fulfil the requirements of the MDC system *i.e.,* the distortion of the joint decoding is less than the distortion of the individual description decoding at required rate.

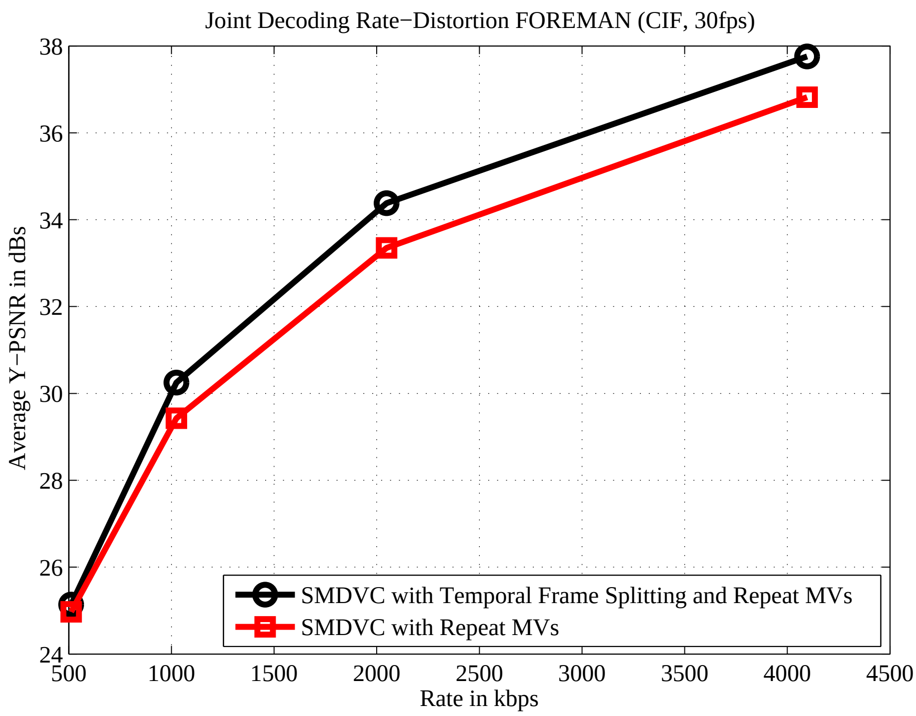
Figure 6.8 shows the block diagram of the SMDVC framework having different multiple motion vector streams. MDSQ is used to create two streams of motion vectors generated from the MCTF block. The step size used for the central quantizer of the MDSQ for motion vector is same as the motion vector accuracy *i.e.,* δ_{MV}. The amount of redundancy between the motion vectors descriptions is controlled by the index assignment matrix parameter f. The decoded motion vector is same as generated at the encoder when both the motion vector streams are available at the decoder. On the other hand, the motion vectors value is changed by $\pm g\delta_{MV}$, when single description is received at the decoder. The effect of creating multiple motion vector streams using MDSQ on joint and individual decoding for different index assignment matrix parameter is presented in Section 6.2.3.

6.2.3 Simulation Results

The simulation results of the proposed SMDVC for different redundancy control schemes are presented in three stages: rate distortion comparison of the proposed SMDVC scheme that considers all frames and high frequency frame splitting, rate distortion performance of the SMDVC when multiple motion vector streams are generated by MDSQ and performance evaluation of different redundancy schemes under a packet erasure channel. The rate distortion performance of joint and individual decoding of the SMDVC scheme by considering all temporal frames and different high pass frames in each description is shown in Figure 6.9 and Figure 6.10 respectively. The redundancy between the descriptions is reduced by splitting high frequency temporal frames in each description. The joint decoding quality by frame interleaving is better than the joint decoding quality having all the temporal frames in each description. The improvement in decoding quality by using frame splitting is due to the availability of more number of refinement levels of each spatio-temporal subband for each description at the decoder at the same data rate. The SMDVC with frame splitting gives on average 0.6 dB better joint decoding PSNR than the scheme that considers all the temporal frames in each description. On the other hand, the individual decoding PSNR of frame splitting is less than the individual decoding PSNR of a scheme that considers all the temporal frames in each description. The decrease in PSNR of the side decoding is due to the unavailability of few high frequency frames in each description.

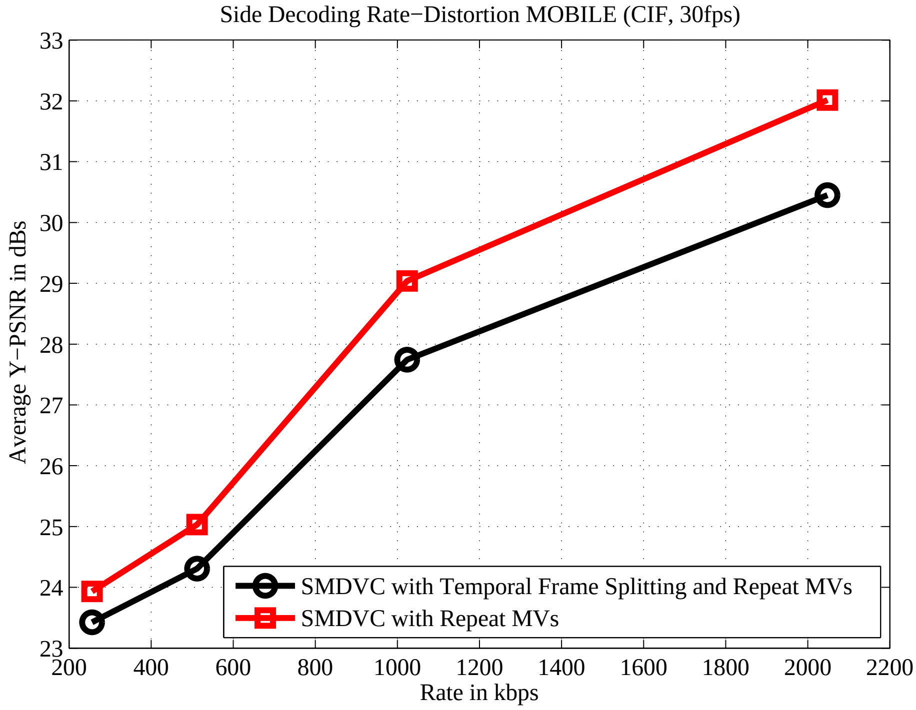


(a)

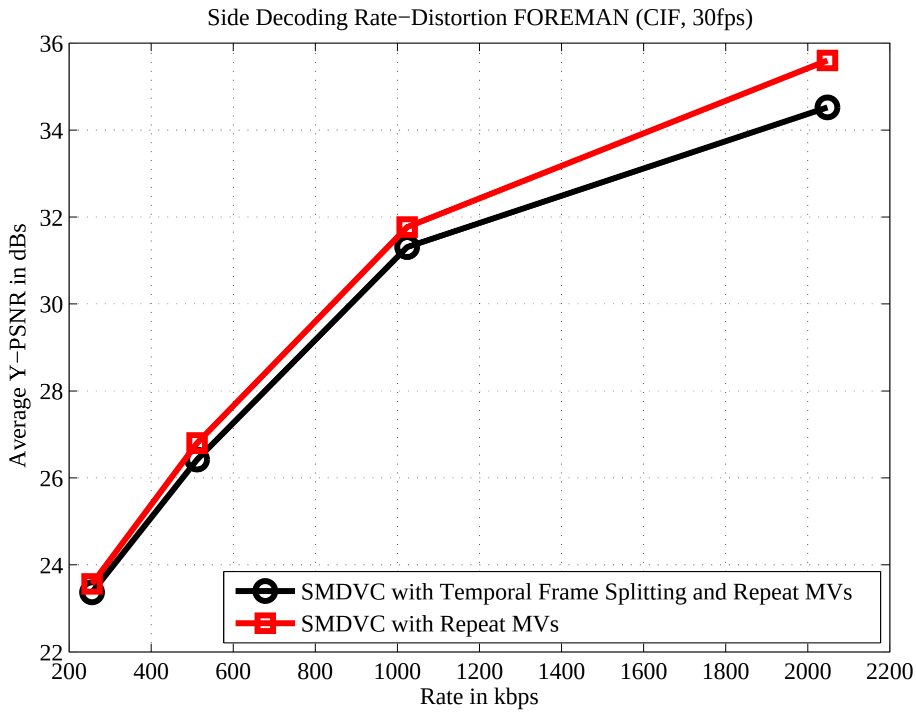


(b)

Figure 6.9: Rate-Distortion comparison of joint decoding of MDSQ-SR-based SMDVC having All frame and high pass frame splitting for (a) *Mobile* (b) *Foreman* sequences.



(a)



(b)

Figure 6.10: Rate-Distortion comparison of individual decoding of MDSQ-SR-based SMDVC having All frame and high pass frame splitting for (a) *Mobile* (b) *Foreman* sequences.

Table 6.1: Motion vector (MV) cost in kbps and joint decoding quality (PSNR in dB) for different sequences.

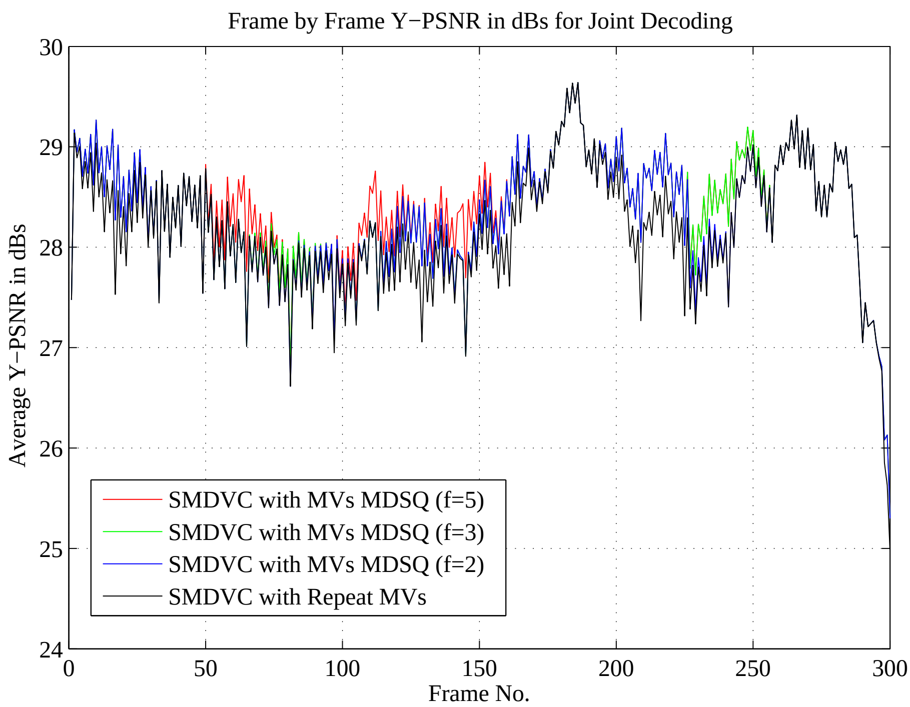
Mobile Sequence (CIF, 30fps)								
Data Rate	Repeated MVs		MVs using MDSQ $f = 2$		MVs using MDSQ $f = 3$		MVs using MDSQ $f = 5$	
R_c in kbps	MVs cost in kbps	Y-PSNR in dBs	MVs cost in kbps	Y-PSNR in dBs	MVs cost in kbps	Y-PSNR in dBs	MVs cost in kbps	Y-PSNR in dBs
400	133.32	26.4262	107.38	26.4756	102.82	26.5256	98.23	26.5402
600	133.32	27.4660	107.38	27.4907	102.82	27.5037	98.23	27.5079
800	133.32	28.1880	107.38	28.3089	102.82	28.3089	98.23	28.4108
Foreman Sequence (CIF, 30fps)								
400	125.39	28.3297	106.49	28.3555	102.92	28.3673	98.65	28.3722
600	125.39	30.2315	106.49	30.3384	102.92	30.5222	98.65	30.6068
800	125.39	31.8655	106.49	31.9092	102.92	31.9180	98.65	31.9242

Table 6.1 shows the motion vector cost and joint decoding PSNR for *Mobile* and *Foreman* sequences for different data rates by using the same motion vector information in each description and by generating two motion vector streams using MDSQ for $f = 2, 3, 5$ in SMDVC framework. It is observed from Table 6.1 that by creating different motion vector descriptions using MDSQ the cost of motion vector information is reduced up to 35kbps and the joint PSNR is improved by 0.22 dBs. Frame by frame Y-PSNR of the joint and individual decoding for *Mobile* sequence using the same motion vector information and motion vector descriptions generated by MDSQ, encoded at 800 kbps is shown in Figure 6.11. It is evident from Figure 6.11 that the side decoding PSNR is rapidly decreased by increasing the number of diagonals filled in the index assignment matrix of the motion vector MDSQ.

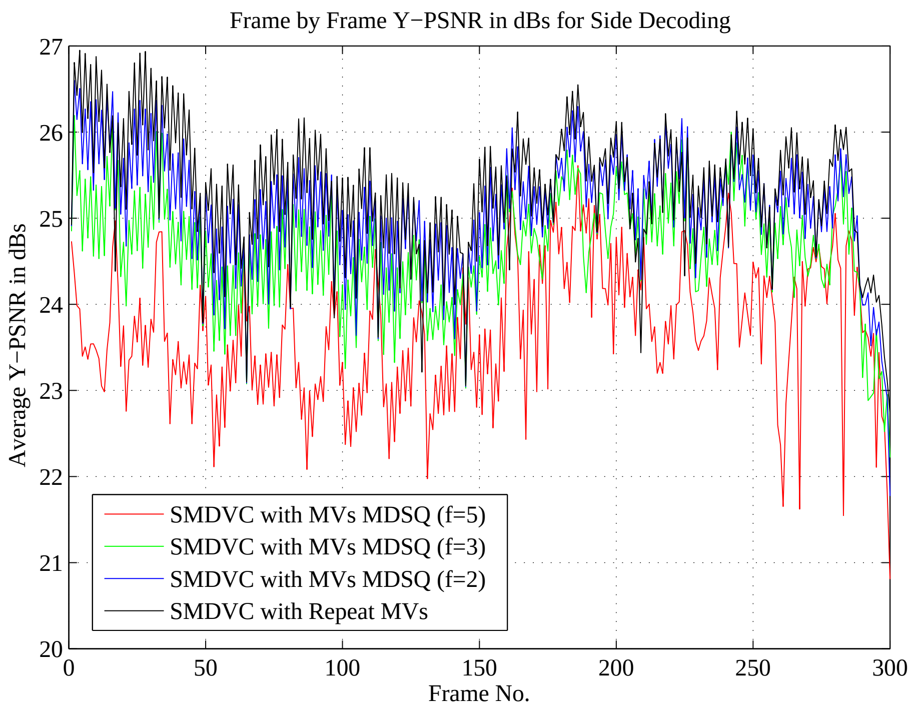
For the performance evaluation over packet erasure channel, the same sort of packetization and error patterns are considered as used in Section 6.1. Figure 6.12 shows the comparison of the MDSQ-SR-based SMDVC considering all frames, high frequency frame splitting with repeat motion vectors and the MDSQ-SR-based SMDVC considering all frames with multiple motion vector streams using MDSQ under packet loss environment for *Mobile* and *Foreman* sequences encoded at 512 kbps. It is observed that by selecting different frames has higher PSNR than all frames at no packet loss. However, the PSNR for high frequency frame splitting is decreased by $0.2 - 0.6$ dBs for percentage packet loss greater than zero. The decrement in PSNR value of joint decoding is due to the missing high pass temporal frames in each description. On the other hand, the performance of SMDVC with motion vector streams from MDSQ is better than repeating the motion vector information for any percentage of packet loss.

6.3 SMDVC Application

In Chapter 5, a fully SMDC scheme for images is presented that uses the concept of MUDC and MDSQ-SR. In this section, both the MUDC and MDSQ-SR techniques are integrated for generating any number of scalable descriptions for video and is evaluated for P2P video streaming as an application for the proposed framework. In P2P video streaming, the robustness to peer and packet losses is

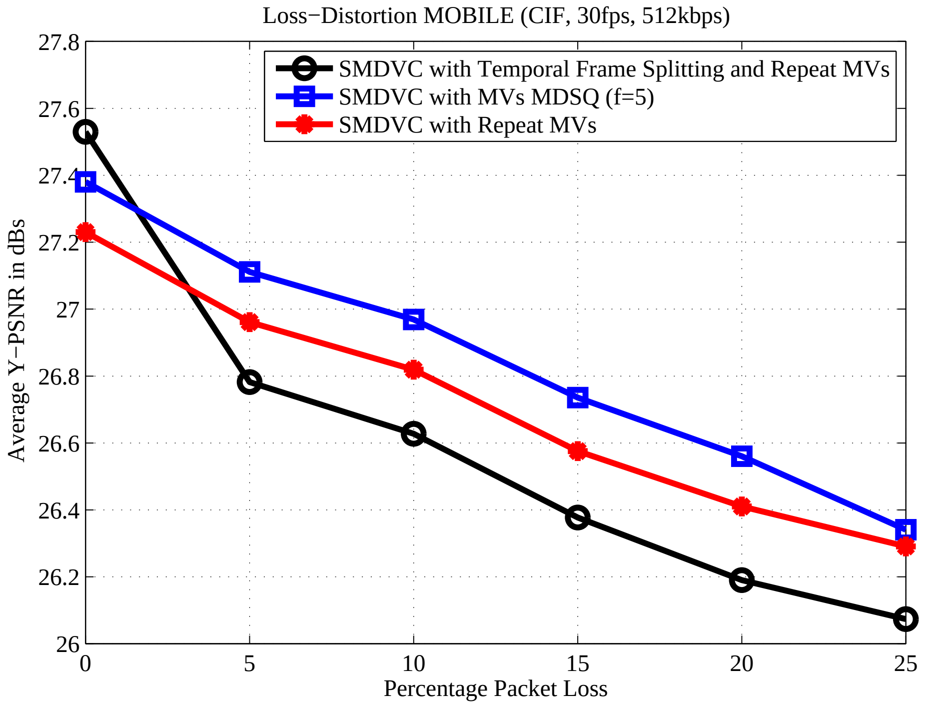


(a)

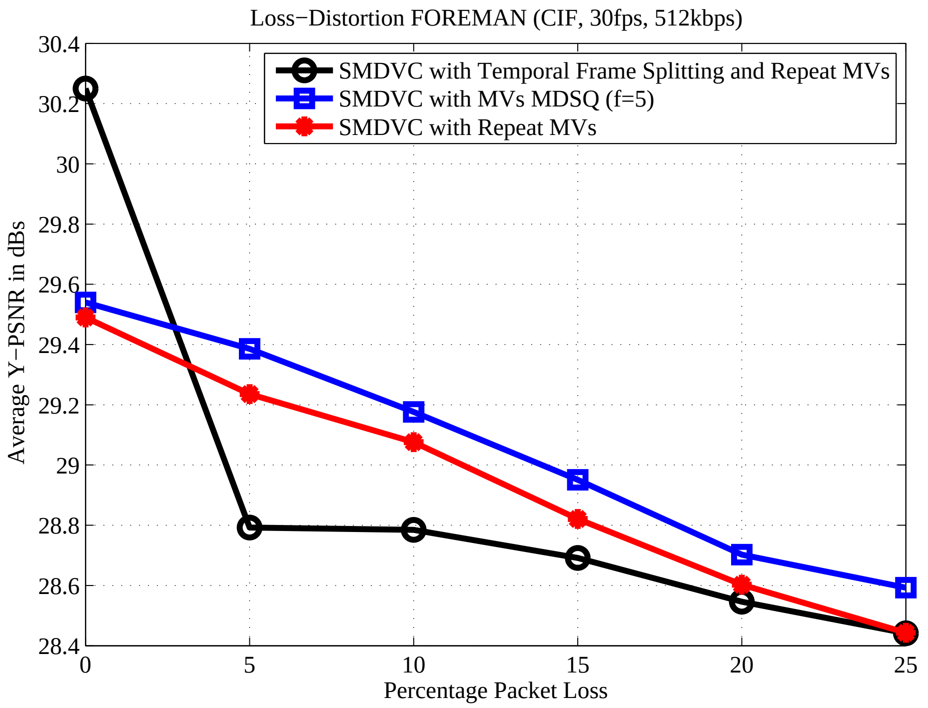


(b)

Figure 6.11: Frame by Frame Y-PSNR of the *Mobile* sequence using repeat MVs and MVs using MDSQ for (a) joint (b) individual decoding.



(a)



(b)

Figure 6.12: Loss-Distortion performance comparison of three different MDSQ-SR based SMDVC schemes for (a) *Mobile* and (b) *Foreman* sequences.

regarded as very important in order to enjoy a good quality of experience. In addition to high robustness, efficient controllability of data rate and redundancy among the data streams from peers is also vital to P2P streaming. In this section, a novel framework for SMDVC scheme for P2P video streaming is presented by addressing these requirements.

6.3.1 P2P Video Streaming

Multimedia communication over Internet has seen a great increase due to success in communication and source coding technologies. In a traditional client/server-based video-on-demand streaming system, multimedia content is stored in one or more dedicated servers and whenever a client requests a particular video, the network redirects the user to one of the servers to stream the required video to the end user. The main problem of such client/server-based systems is the cost, as each server requires large storage space, high bandwidth and the performance reliability. In addition, the conventional client/server-based systems are not scalable and limit the number of users served at one time for high bandwidth applications. P2P networks have emerged as an alternative approach for streaming video over Internet. In P2P networks, there is no need for dedicated server to store and stream the multimedia content because a peer is an ordinary computer that can act as a client and server simultaneously. Each peer has certain storage space and uplink bandwidth for streaming to other clients. P2P networks provide a scalable solution for multimedia streaming. They are dynamic in nature where the peers can leave and join the network, accordingly.

Besides the low cost and scalability properties of the P2P networks, there are certain issues related to P2P streaming that need to be addressed while providing any solution. In P2P networks, the upstream rate of each peer is much less than the downstream rate. For high quality video streaming, the required video data rate can exceed the uplink bandwidth of peers. Therefore, distributed video streams are required for each peer to cater for the uplink bandwidth problem for high quality video streaming that also provides balancing of the load among peers. Each peer in P2P networks is an ordinary computer and connected to the Internet through different speed connections, therefore peers should be able to adapt the streams according to the data rate requirements, while minimizing the effect on the visual quality. Due to heterogeneous nature of network conditions for peers and the use of distributed video streams among different peers, a mechanism for peer selection is required in such schemes. Similarly, failure recovery mechanisms are also required for P2P networks, as not all peers are active at the time. Different methods are available in the literature for video streaming over P2P networks that consider how to distribute multimedia content among different peers [109, 110], how to adapt the streams according to network conditions [111, 112] and how to select different peers and recover when peers are not active [113, 114].

During video streaming over P2P network, any peer can leave the network, links to any peer can be broken and packets from any peer can be lost. It is difficult for the receiving peer to get acceptable video quality if the base layer link is broken or packets from base layer are lost. MDC provides an effective solution that overcomes the problem of data loss in such streams [115–119]. For multiple descriptions generation, some amount of redundancy is added in streams leading to graceful degradation in video quality, if peers are turned off or packets from any stream are lost. Different MDC solutions for P2P networks have been proposed based on different video coding standards (*e.g.,* (MPEG-4 [120, 121], H.264/SVC [122] and MCTF+JPEG2000 [123, 124]), different methods for descriptions creation (*e.g.,* multiple description scalar quantizer (MDSQ) [125], spatial-temporal subsampling [121] and coupled with forward error corrections [120]) and different applications like IP-TV [126] and communication over mobile links [127].

In [121], two descriptions are generated by temporally subsampling and followed by encoding with standard MPEG-4 codec, and embedding the quarter sized subsampled coded streams temporally to increase the reliability. At the receiving peer, video can be decoded by either temporal, spatial or hybrid spatial-temporal interpolation depending on which and how many descriptions are received. In [117], wavelet-based pre and post processing MDC is proposed in which generated descriptions are not equally important and is assumed that the most important description is transmitted over a reliable path. In another approach, flexible multiple descriptions are created from a single scalable bitstream generated from MCTF+JPEG2000 based scalable video codec [115]. The generated flexible scalable multiple descriptions are then adapted depending on network

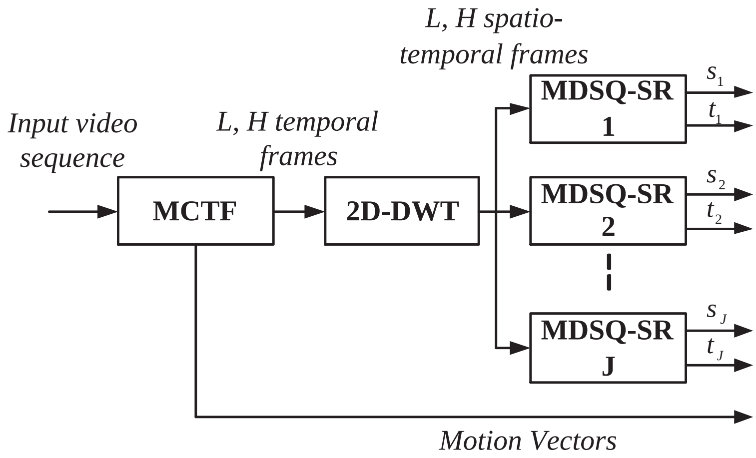


Figure 6.13: Block diagram of SMDVC encoder for P2P Streaming.

conditions by varying the number of base and enhancement layers, redundancy and rate of each description. In [124], it is shown that the received video quality is improved if both the redundancy and rate are adapted simultaneously on fly.

In this section, a novel approach for adapting the redundancy and data rate of descriptions coming from different peers is proposed. The constrained successive refinement of MDSQ is used to obtain the robust scalable multiple descriptions and then the MUDC technique is used for generating and joint decoding of the streams coming from different peers leading to efficient control of redundancy and data rate. The main contribution of combining the MDSQ-SR and MUDC is two-fold:

1. The use of MDSQ-SR in MDC environment enables each description to provide robust quality scalability in individual and joint decoding.

2. The use of MUDC facilitates the peers to send descriptions of different bandwidth availabilities of the peers resulting in better joint decoding quality.

6.3.2 SMDVC for P2P Streaming

The multiple description video coding scheme proposed for P2P streaming is based on MCTF, MDSQ-SR and MUDC. The block diagram of the proposed

SMDVC encoder for P2P streaming is shown in Figure 6.13. Firstly, MCTF is performed on an input sequence to remove the temporal correlation among the frames. After temporal decomposition, each temporal frame is spatially decomposed by using 2D-DWT to remove the correlation within the frame. MCTF and 2D-DWT completes the spatio-temporal decomposition of the input sequence. For generating two scalable descriptions, s and t, the spatio-temporal coefficients are quantized by MDSQ-SR for P refinement levels followed by entropy coding. Similarly, for generating more than two descriptions several MDSQ-SR are used with different base layer by using different central quantizer bin widths. It is assumed that the motion vector information is transmitted securely and embedded in each description, therefore available at the decoder without any losses. Different descriptions from several MDSQ-SRs is selected according to the design constraints of the MUDC as derived in Section 3.2.2 and the design conditions are $a > 1$, $a^{k-j}g_j > g_k$ and $a^{k-j} \geq g_k$, where $k > j$. The amount of redundancy among the description from single MDSQ-SR is selected according to the deign parameters of MDSQ-SR *i.e.,* g and r. Similarly, the joint and individual distortion constraints are satisfied by fulfilling the MDSQ-SR design conditions derived in Section 4.2.2 and the conditions are $r > 1$, $r > g$, and r and g are not integer multiple of each other.

Once the scalable descriptions s and t are encoded from each MDSQ-SR, each peer extracts only the subset of temporal levels, spatial subbands and refinement levels, p, of the MDSQ-SR according to the frame rate, resolution, quality and data rate preferences of the requesting peer. The extracted descriptions at each peer are selected in such a way to minimize the distortion at the uplink bandwidth of the sending peer. In the proposed scheme, rate is allocated to each description by considering all the frames and spatial levels, but different refinement levels p for each temporal and spatial subbands provided that distortion is minimum at the given bit budget. Finally, the decoder at the receiving peer is capable of decoding the bitstreams either individually or jointly depending on how many peers are active, *i.e.,* how many descriptions are received.

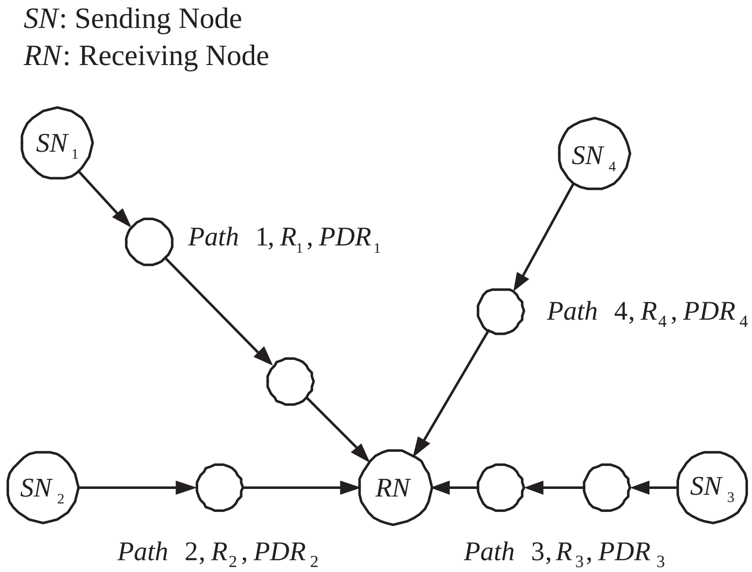


Figure 6.14: Simulation setup for P2P streaming.

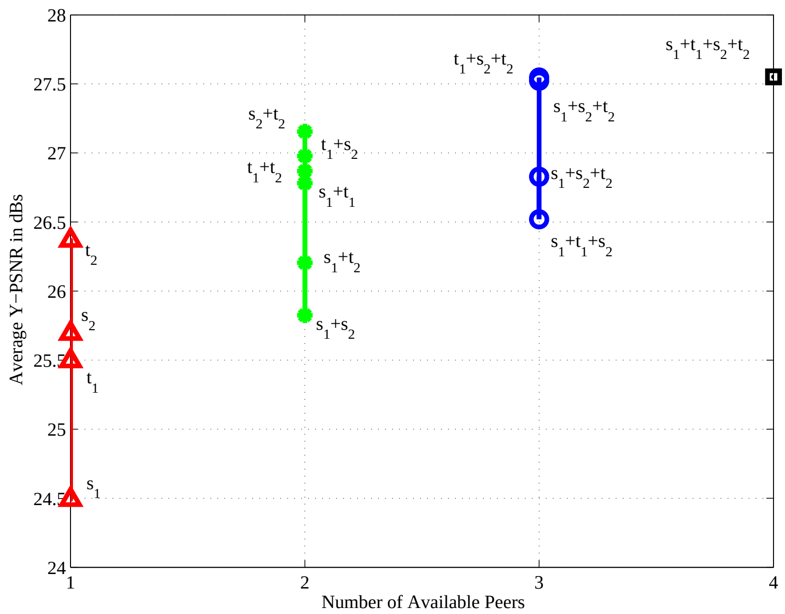
6.3.3 Performance Evaluation of SMDVC for P2P Streaming

For the simulations, $N = 4$ serving peers and one receiving peer is considered and are connected to each other as shown in Figure 6.14. Each sending peer has description generated from the SMDVC framework presented in Section 6.1. For the initial simulations, a homogenous network conditions are assumed, where each sending peers has the same uplink bandwidth, *i.e.,* $R_1 = R_2 = R_3 = R_4$. The source coding rate for each extracted description is set to 512 kbps and the download bandwidth of the receiving peer is set to 2 Mbps. The heterogeneous network conditions are also considered to demonstrate the efficiency of the scalability of SMDVC bitstreams. As shown in Figure 6.14 each sending node is connected to the receiving node through different paths. The packet drop rate (PDR) for each independent path between the sending and receiving peer varies between 0% and 25% during the whole streaming session. It is also assumed that each sending peer is either in an active or non-active state. The active and non-active state means that a peer is available or not available for streaming the video to the receiving peer. The peer available time for streaming is modelled by exponentially distributed random variable as in [109] and [120]. Similarly, the time for which peer is not available for streaming is also modelled by another exponentially distributed random variable.

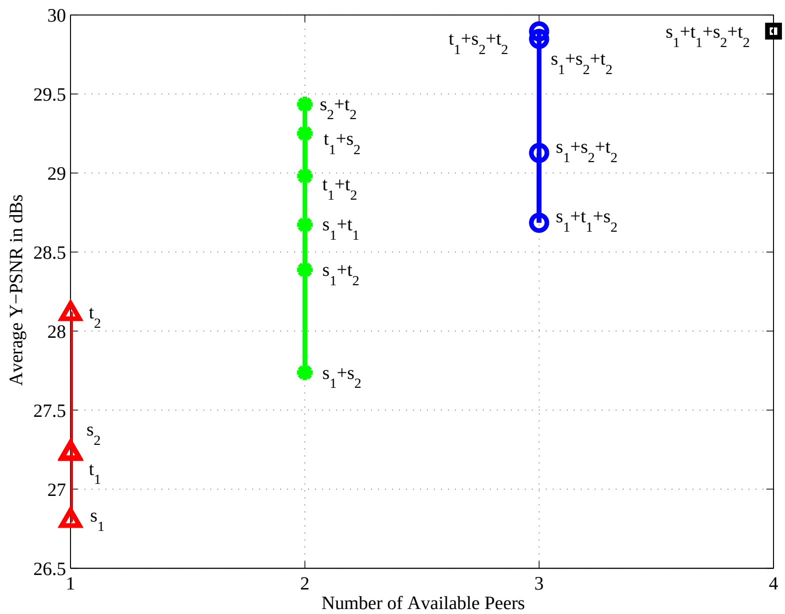
Each peer corresponds to a different description generated from two different MDSQ-SRs. Let s_1, t_1 be the two scalable descriptions created from MDSQ-SR 1 and s_2, t_2 are the remaining two scalable descriptions created from the MDSQ-SR 2. The base layers of the two MDSQ-SRs follow the multichannel unbalanced description generation, for example as in Figure 3.4. The two MDSQ-SRs are related to each other by the factor $a = 3$. Each quality scalable description is created from MDSQ-SR after 4 levels of temporal and 5 levels of spatial decomposition. The number of successive refinement levels for each MDSQ-SR is $p = 3$. The parameters related to each MDSQ-SR are $f = 2$ and $r = 3$, where f is the number of diagonals filled in the index assignment matrix and r is the successive refinement factor of each side quantizer bin. These parameters are selected based on the conditions proposed in Chapter 3 and Chapter 4 respectively. The performance of the proposed SMDVC scheme for P2P streaming is presented in three stages: performance under lossless conditions, performance evaluation under packet erasure channel with fixed number of available peers and performance under packet erasure channel with peers active and non-active periods.

6.3.3.1 Performance Evaluation under Lossless Conditions

Figure 6.15 shows the PSNR values of the decoded *Mobile* and *Foreman* sequences from different combinations of available peers *i.e.,* $N = 1, 2, 3, 4$. It is observed from Figure 6.15 that few combinations give large PSNR improvement than other combination due to different side quantizer bin spread of the base layer of different MDSQ-SR. For $N = 4$ peers the decoded PSNR is the same as one of the decoded PSNR from $N = 3$ peers. Same kind of improvement can also be observed from the SSIM and VQM results presented in Figure 6.16 and Figure 6.17 respectively. The advantage of using these peers is on increasing the robustness of the system as demonstrated in the following sub sections.

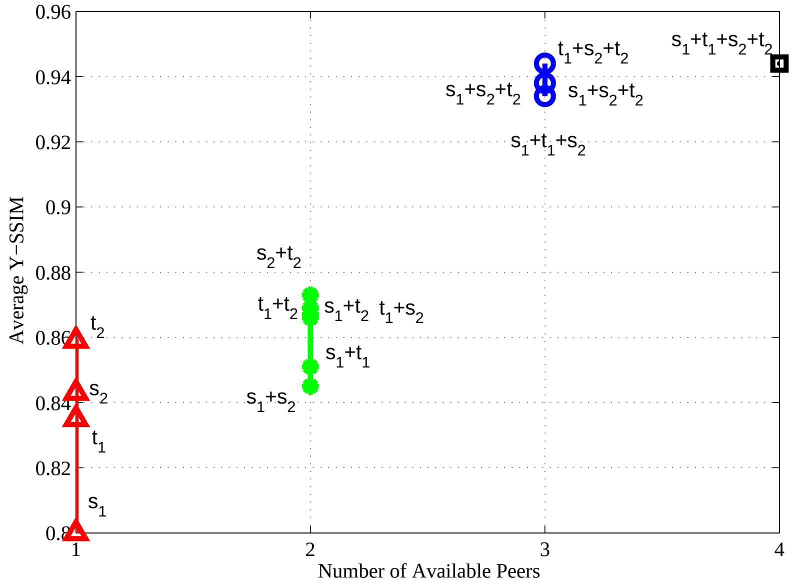


(a)

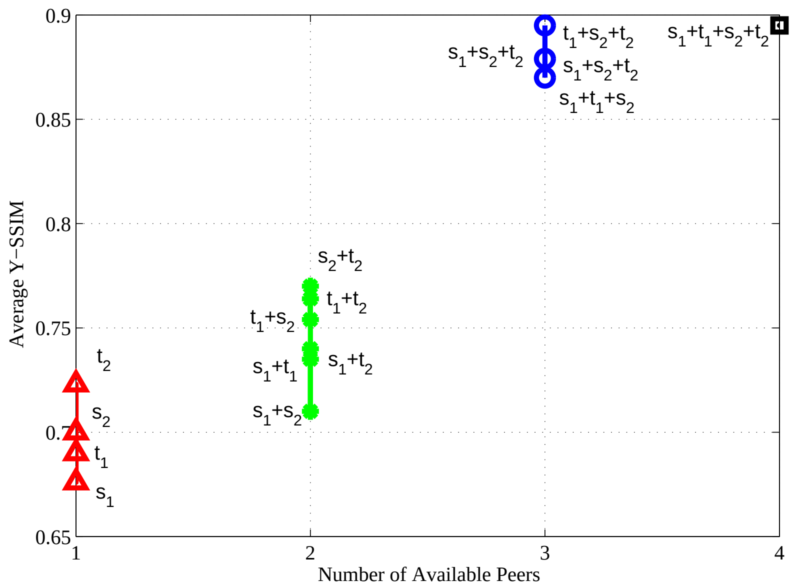


(b)

Figure 6.15: Decoded Y-PSNR from different combinations of available peers for (a) *Mobile* (b) *Foreman* sequence.



(a)



(b)

Figure 6.16: Decoded Y-SSIM from different combinations of available peers for (a) *Mobile* (b) *Foreman* sequence.

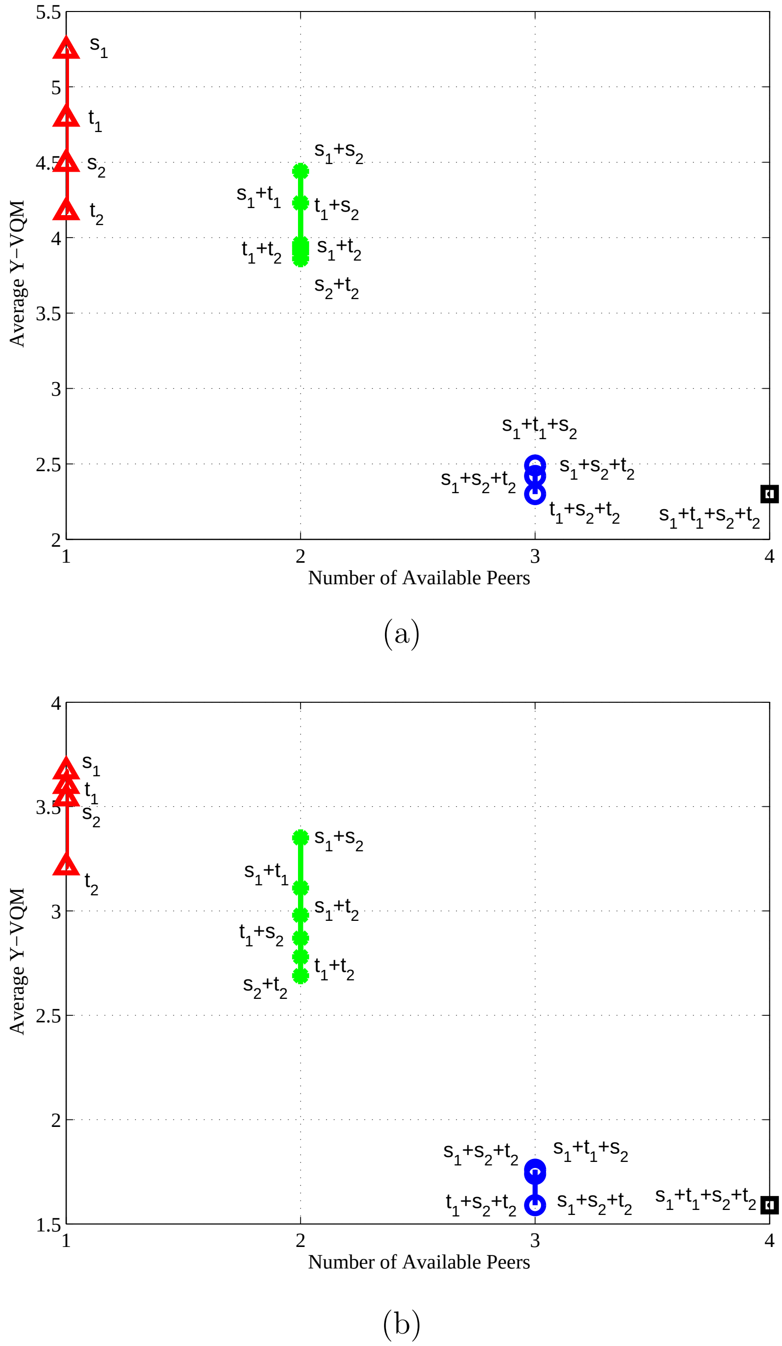


Figure 6.17: Decoded Y-VQM from different combinations of available peers for (a) *Mobile* (b) *Foreman* sequence.

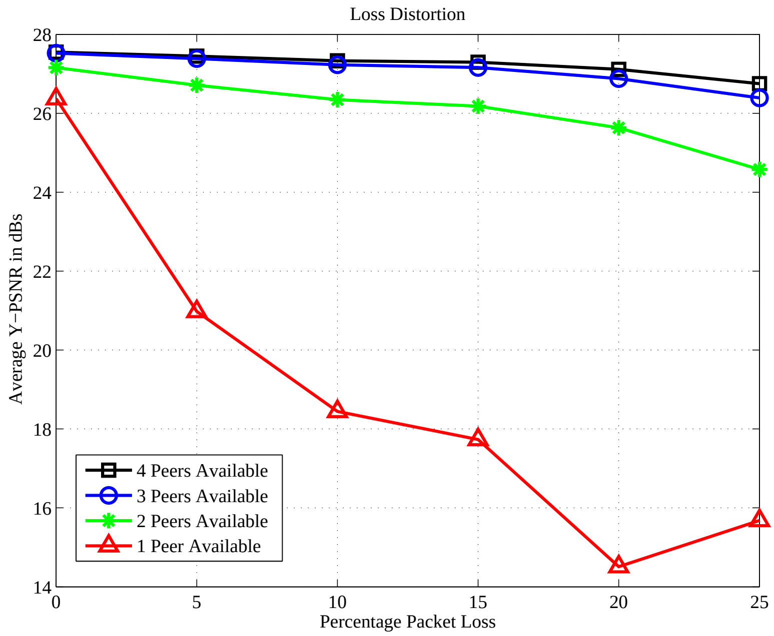
6.3.3.2 Performance Evaluation under Packet Erasure Channel with a Fixed Number of Available Peers

The goal of this set of simulations is to evaluate the effect of packet loss on the decoded quality for different number of available peers. For this purpose, each extracted scalable description is packetized and different packet loss percentages are considered for each description. The coefficients of each temporal and spatial refinement levels are placed in the same packet to minimize the erasure effect. Let M be the total number of packets for each description and l be the number of lost packets. Then there is a total of ${}^{M}C_{l}$ number of combinations to loose l packets from the total of M packets. The average PSNR for a particular number of packet losses is calculated by averaging the PSNR for a different number of combinations.

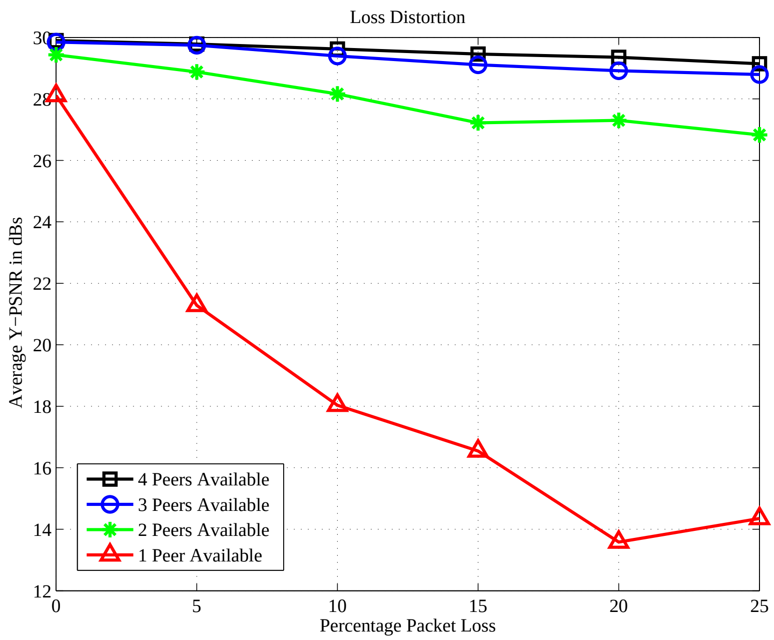
For this set of experiments, the packet erasure channel model having the loss rate varying between 0% and 25% is considered. Figure 6.18 shows the average PSNR of the decoded sequence from $N = 1$ to $N = 4$ available peers having different percentage of packet losses per P2P path. It is evident that the PSNR drop is much smaller for more number of peers available even if the percentage packet loss per path is high. The more robustness behaviour of joint decoding of more number of descriptions under packet erasure channel can also be observed from SSIM and VQM results shown in Figure 6.19 and Figure 6.20 respectively. Figure 6.21 and Figure 6.22 shows the GOP by GOP PSNR comparison of the best possible combinations from $N = 1$ to $N = 4$ available peers having (a) 0%, (b) 5%, (c) 10%, (d) 15%, (e) 20%, and (f) 25% packet loss per P2P path. The superior visual quality for higher number of available peers can be observed in Figure 6.23, from the portion of the decoded *Mobile* sequence after 25% packet loss rate in each path.

6.3.3.3 Performance Evaluation under Varying Packet Loss Rate and Varying Number of Available Peers

The main aim in this section is to evaluate the proposed SMDVC scheme over a P2P network as shown in Figure 6.14. To evaluate the decoded quality of the

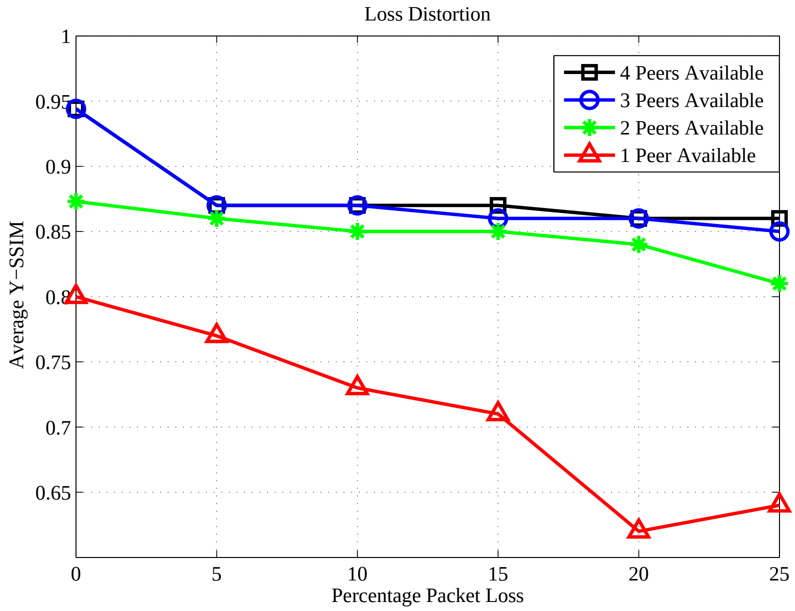


(a)

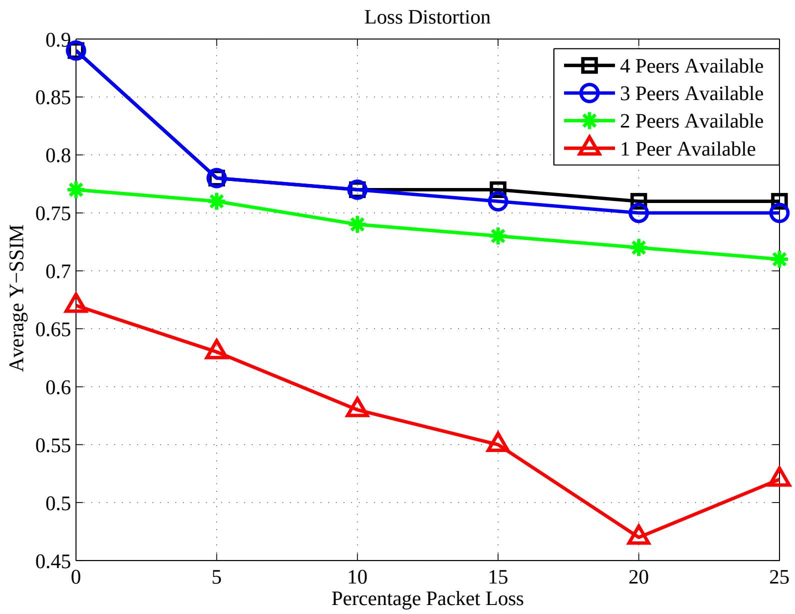


(b)

Figure 6.18: Loss Distortion performance in terms of Y-PSNR from different combinations of available peers for (a) *Mobile* (b) *Foreman* sequence.

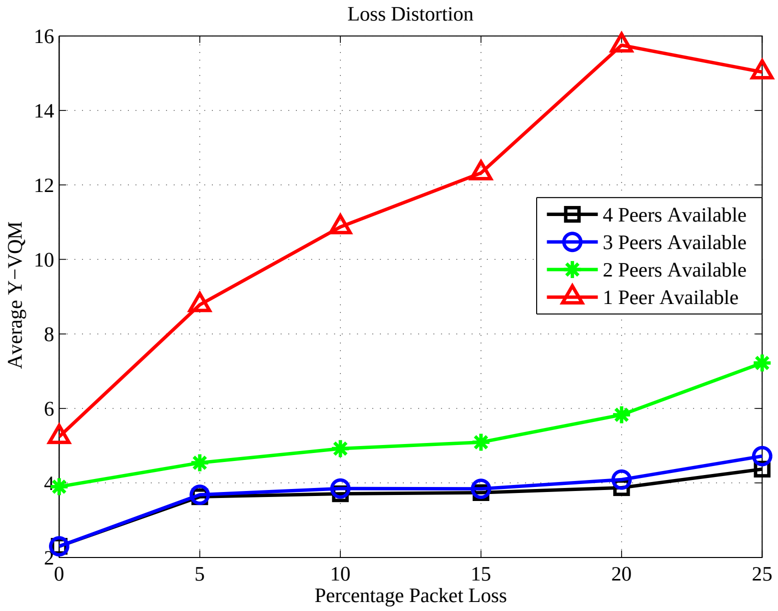


(a)

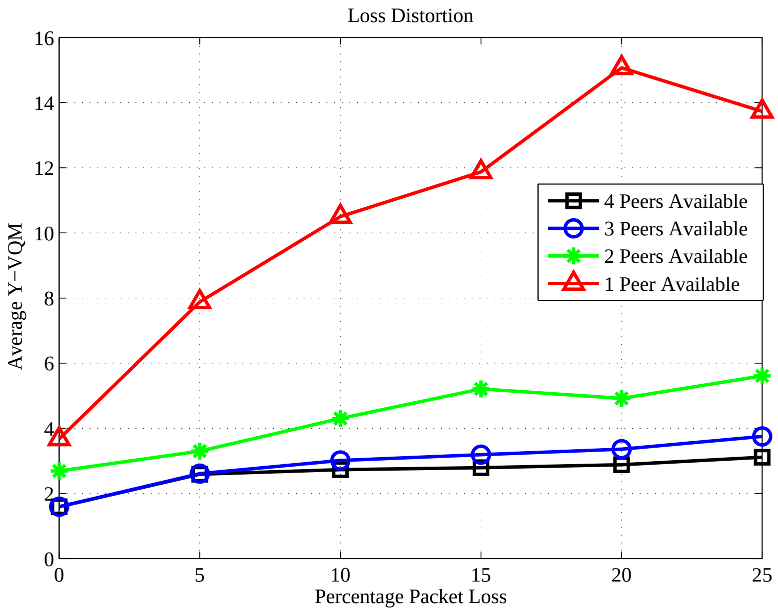


(b)

Figure 6.19: Loss Distortion performance in terms of Y-SSIM from different combinations of available peers for (a) *Mobile* (b) *Foreman* sequence.



(a)



(b)

Figure 6.20: Loss Distortion performance in terms of Y-VQM from different combinations of available peers for (a) *Mobile* (b) *Foreman* sequence.

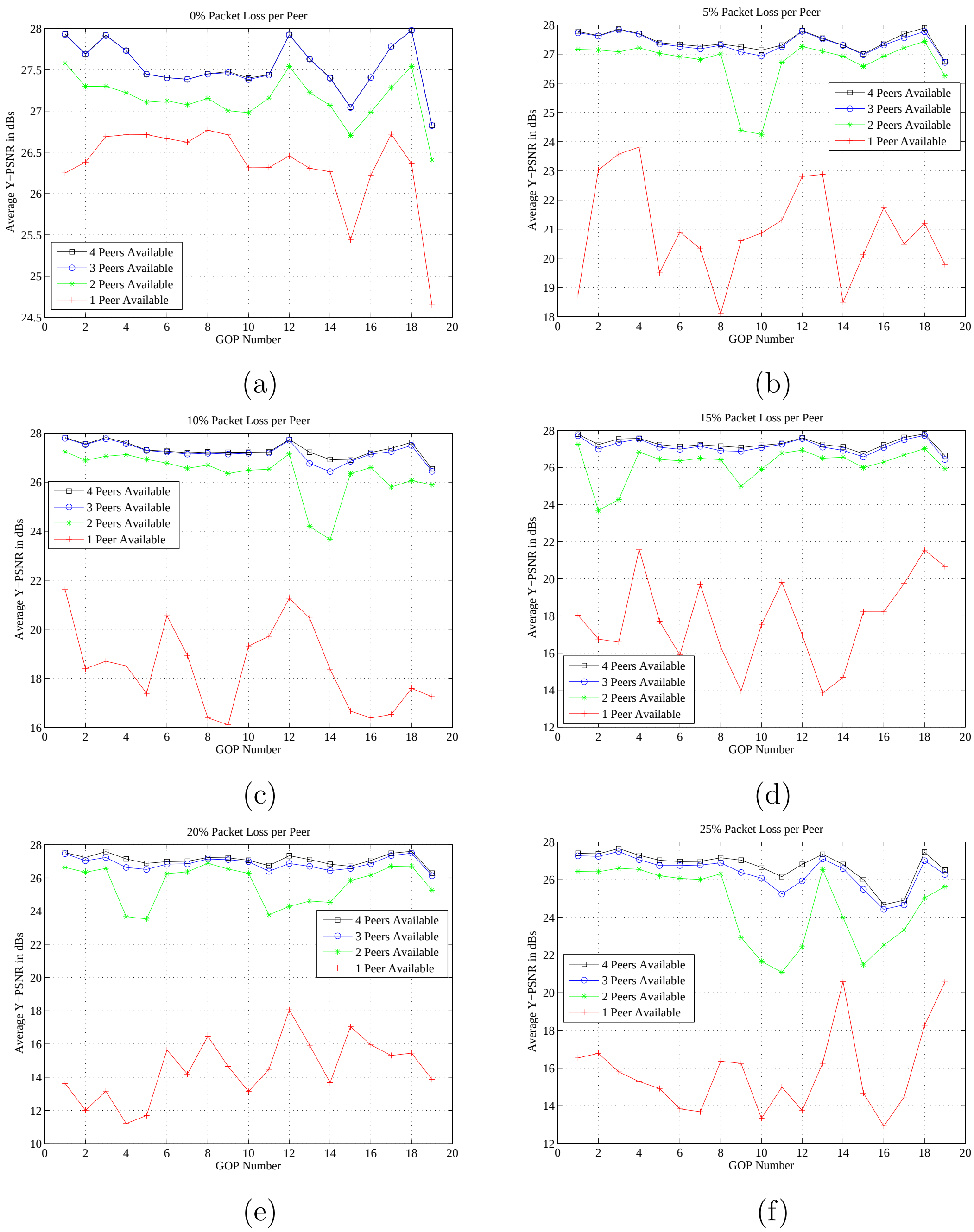


Figure 6.21: GOP by GOP Y-PSNR comparison of the decoded *Mobile* sequence after different percentage of packet loss (a) 0% (b) 5% (c) 10% (d) 15% (e) 20% and (f) 25%.

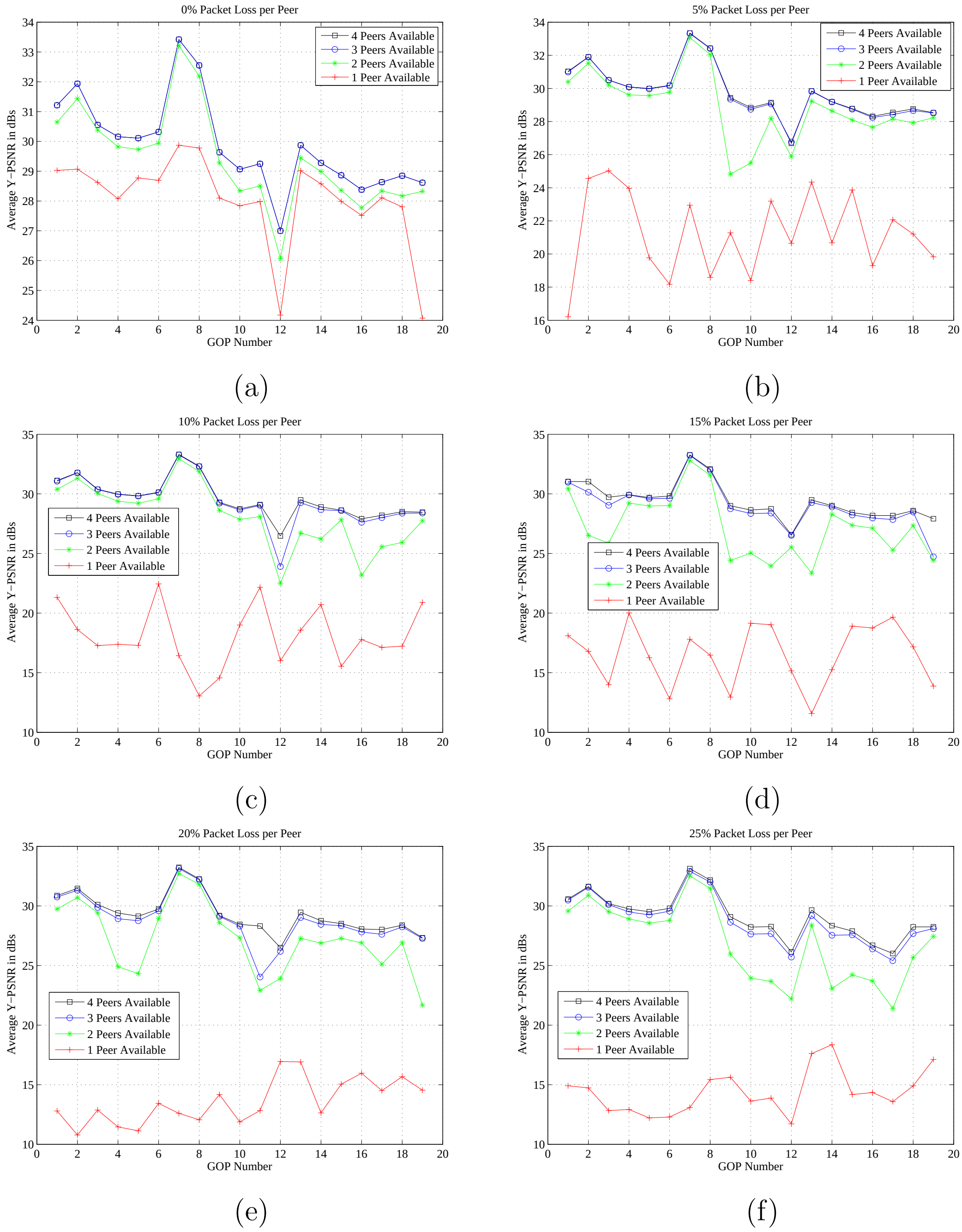


Figure 6.22: GOP by GOP Y-PSNR comparison of the decoded *Foreman* sequence after different percentage of packet loss (a) 0% (b) 5% (c) 10% (d) 15% (e) 20% and (f) 25%.

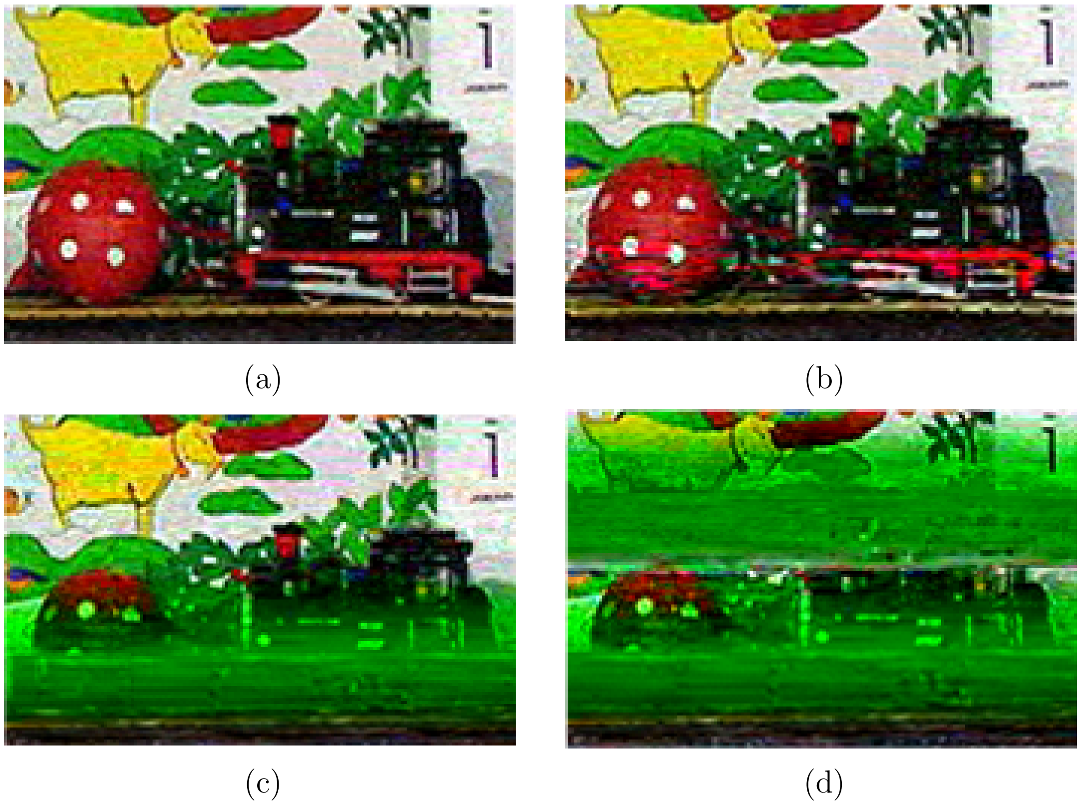


Figure 6.23: Portion of the decoded frame 160 of *Mobile* sequence after 25% packet loss in each path when (a) 4 (b) 3 (c) 2 and (d) 1 peers are available.

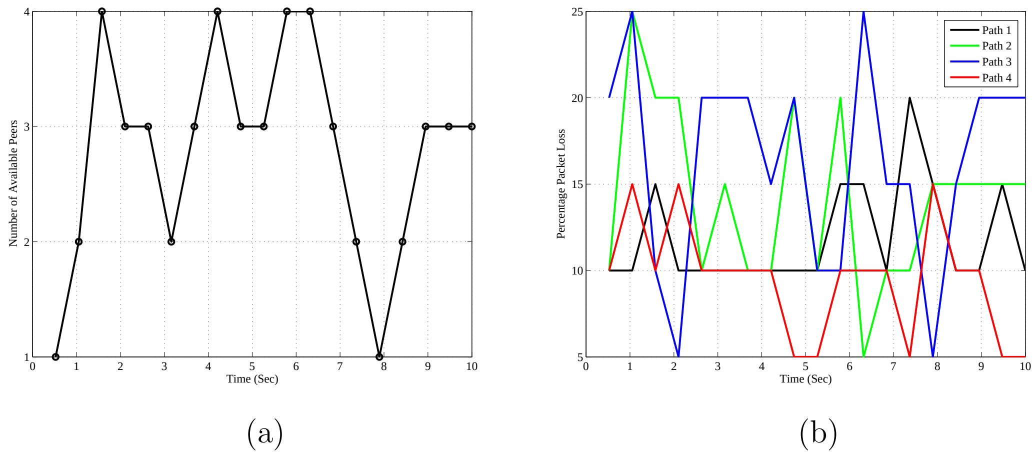


Figure 6.24: Performance evaluation setup for varying conditions (a) Number of available peers (b) Percentage packet loss per path with respect to time.

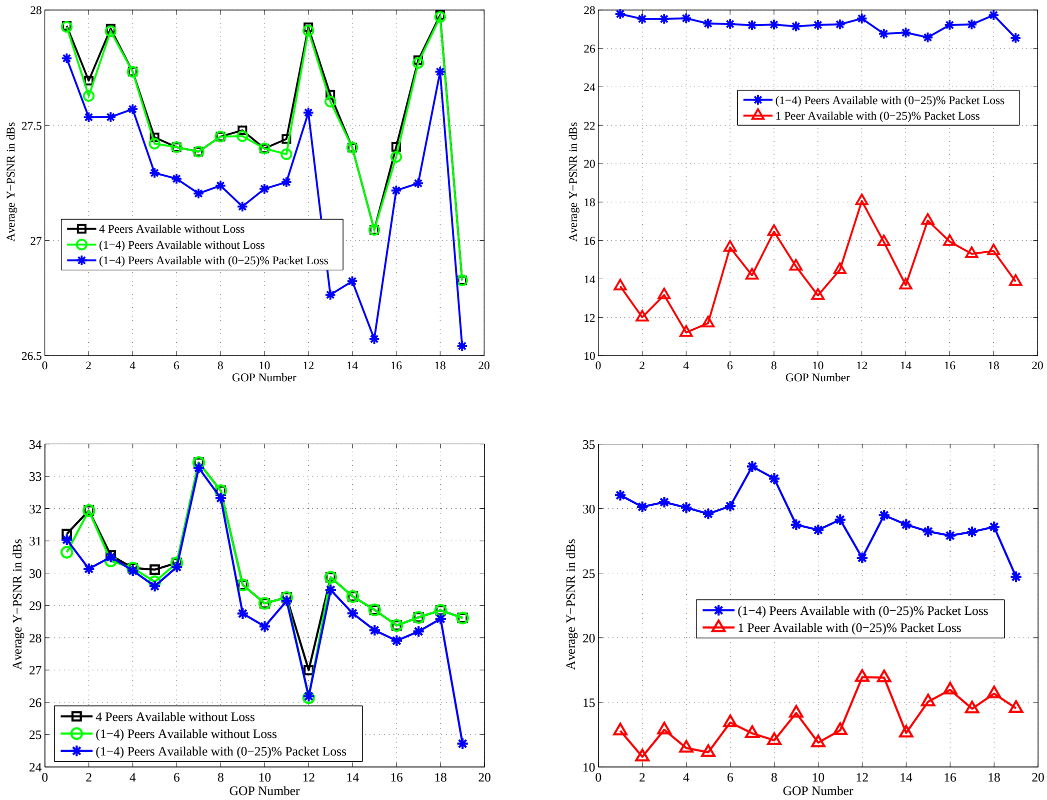
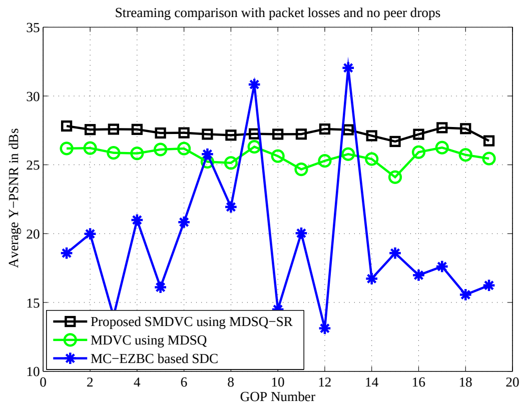


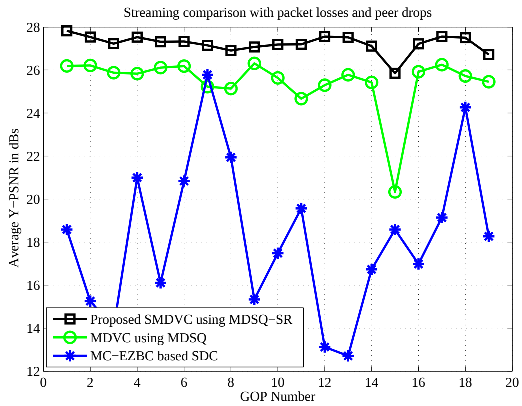
Figure 6.25: GOP by GOP Y-PSNR comparison of the decoded *Mobile* (top row) and *Foreman* (bottom row) sequence under varying channel conditions with 4 peers available (left column) and single peer available (right column).

(a) (b) (c)

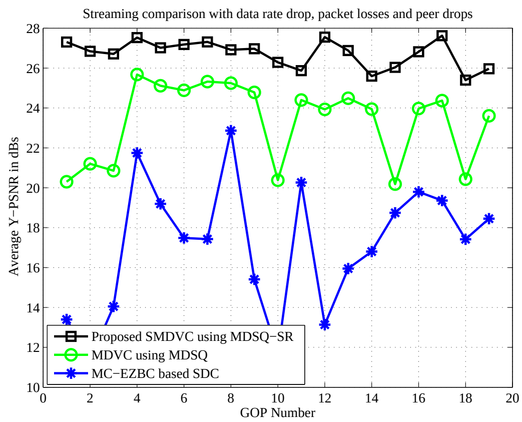
Figure 6.26: Portion of the decoded frame 12 from GOP 11 of *Mobile* sequence when (a) 1-4 peers available with no loss (b) 1-4 peers available with 0-25% packet loss (c) 1 peer available with 0-25% packet loss.



(a)



(b)



(c)

Figure 6.27: GOP by GOP Y-PSNR comparison of the decoded *Mobile* sequence for three different scenarios (a) $(0-25)\%$ packet losses without peer drops outs (b) $(0-25)\%$ packet losses with peer drops outs (c) $(0-25)\%$ packet losses with peer drops outs and data rate reduction.

video over P2P network, when random packet loss patterns for each active path is considered. Performance under packet erasure channel is evaluated. The simulation is performed by randomly choosing 20 different packet loss patterns for each path from the total packet loss patterns as generated in Section 3.4.2 and then the final result is formed by averaging all the individual results. Figure 6.24 shows how many peers are in active or non-active state and how the packet loss rate changes with time for each path throughout the streaming session. Figure 6.25 shows the GOP by GOP PSNR comparison of the proposed SMDVC scheme for P2P network streaming under different packet drop rates and different number of available peers. Figure 6.26 shows the portion of decoded frame of *Mobile* sequence under varying channel conditions when (a) 1-4 Peers Available without Loss (b) 1-4 Peers Available with 0-25% Packet Loss (c) 1 Peer Available with 0-25% Packet Loss. The corresponding superior visual quality for higher number of available peers is evident in Figure 6.26, from the portion of a decoded frame from *Mobile* sequence with up to 4 peers available and with up to 25% packet loss rate in each path.

Figure 6.27 shows the comparison of the performance of the proposed SMDVC scheme that combines MUDC and MDSQ-SR with that of multiple description video coding that uses MDSQ (which is the state-of-the-art MDC method) and that of spatio-temporal decomposition based single description scalable coding for P2P streaming session for the network topology shown in Figure 6.14. For the proposed SMDVC method and MDSQ- based MDVC scheme the description considered from each peer is independent of each other. For single description scalable coding, the base layer is repeated in each peer while each peer has different enhancement layers. Figure 6.27 shows the PSNR of the Y component measured for GOP by GOP of the decoded *Mobile* sequence (CIF, 30fps, 512kbps) simulated under three different scenarios: (a) Between 0%-25% varying packet losses in each path without peer drop outs in a homogenous P2P network conditions; (b) Between 0%-25% varying packet losses in each path with peer drop outs in a homogenous P2P network conditions; (c) Between 0%-25% varying packet losses in each path with peer drop outs in a heterogeneous P2P network conditions with data rate from two of the nodes becoming 250 kbps. Since all four streams are source coded at 512 kbps, the last scenario requires dropping of half of the original bit streams. This scenario corresponds to the four descriptions generated from example (c) and (d) in Figure 4.2. In the first scenario, the proposed SMDVC scheme shows PSNR improvement of 1.69 dB and 7.38 dB with respect to the MDSQ-based MDVC and single description scalable coding, respectively. In the second scenario, the corresponding improvements are 1.77 dB and 9.03 dB, respectively. For the final scenario, these improvements rise to 3.4 dB and 9.64 dB, respectively. These results demonstrates the superiority of the proposed SMDVC scheme, which shows clear benefit of making multiple description bitstreams scalable in P2P streaming under both homogenous and heterogeneous network conditions.

6.4 Summary

In this chapter, we proposed a new MCTF-based framework for SMDVC using MDSQ-SR and MUDC. MDSQ-SR was used to provide the quality scalability in each description, while the temporal decomposition was used to decode the video at different frame rate. The method for selecting different spatio-temporal subbands was also presented for rate allocation of each description in such a way that the joint decoding gives the minimum distortion. Rate distortion performance of the individual and joint decoding of the MDSQ-SR-based SMDVC is presented by truncating each description at different data rates. It is evident from the simulation results that the joint decoding of the MDSQ-SR-based SMDVC gives 0.3 dB better performance than EMDSQ-based MDVC having same amount of redundancy. The loss distortion of the MDSQ-SR-based SMDVC outperforms the MC-EZBC-based single description coding when compared under packet erasure channel.

The problem of inter description redundancy control in SMDVC was also addressed in this chapter and two different schemes were proposed. The texture information redundancy is reduced in one of the proposed scheme, while the motion vector redundancy is reduced in the other scheme. The texture redundancy in the proposed SMDVC is reduced by splitting the temporal high frequency frames in different descriptions. The joint decoding of SMDVC with temporal frame splitting has shown 0.6 dB better performance than the joint decoding of the SMDVC with all temporal frames in each description. It is observed that by selecting different frames has higher PSNR than all frames at no packet loss.

However, the average PSNR for high frequency frame splitting is decreased by $0.2 - 0.6$ dBs for percentage packet loss greater than zero. The motion vector redundancy is reduced by generating multiple motion vector streams using MDSQ instead of repeating motion vector information in each description. The amount of redundancy between the descriptions is reduced by temporal frame splitting at the cost of sacrifice in robustness performance. However, the multiple streams of the motion vector information generated from MDSQ have resulted in reduction of the data rate and robustness enhancement when compared with frame splitting method.

Finally, the integration of MDSQ-SR and MUDC for SMDVC was evaluated for P2P video streaming. By integrating these two techniques, different scalable descriptions were extracted from each peer according to the peer uplink bandwidth, which resulted in the joint decoding at the receiving peer from the same MDSQ-SR and different MDSQ-SR at similar and different data rates. By generating multiple scalable descriptions results in high robustness to packet losses and peer losses (*i.e.,* the active or non-active times) as well as the network bandwidth reductions due to bottleneck links in the network. The comparison of the proposed scheme has shown significant improvements over conventional MDSQ-based MDVC and over single scalable description video coding.

Chapter 7

Conclusions

In this thesis, research on scalable multiple description image and video coding was presented that provides quality, resolution and frame rate scalability to MDC system. The first contribution of this work is to provide an MDC method for any arbitrary number of descriptions that is capable of encoding and jointly decoding any number of descriptions in balanced and unbalanced manner. The second contribution of this research is on creating quality scalable description by using the concept of successive refinement in the side quantizers of MDSQ. Finally, the scheme for generating any number of scalable descriptions is integrated in MCTF-based video coding system to provide fully SMDVC framework, which is also evaluated for P2P video streaming.

In Chapter 3, a novel scheme for generating any number of descriptions using several MDSQs was proposed that provides balanced and unbalanced joint decoding. Hierarchically defined MDSQs that use the successive refinement of the central quantizers were used in the proposed framework. The main parameters include the central quantizer refinement factor, a and the index assignment matrix parameter g_m for the m^{th} MDSQ in the hierarchy. The joint decoding distortion constraints for different combination of descriptions from several MDSQs were formulated with the objective to improve the distortion as the number of jointly decoded descriptions is increased. For meeting these constraints, the design conditions for several MDSQs were proposed. The sufficient and necessary conditions for meeting the required rate-distortion constraints from two MDSQs, $m = j$ and

$m = k$, where $k > j$, are $a > 1$, $a^{k-j}g_j > g_k$ and $a^{k-j} \geq g_k$. These conditions were verified for different combination of parameter values, demonstrating the PSNR increment. It was observed that at the same rate, the unbalanced joint decoding gives 1.1 dB better performance than the balanced joint decoding. An efficient realization of the MUDC scheme was also shown by proposing the successive side quantizer bin merging of the initial MDSQ. It was evident that the rate distortion performance for joint decoding of descriptions coming from different MDSQs resulting in superior performance compared to the balanced descriptions coming from the same MDSQ. The flexibility to add and remove redundancy in terms of number of descriptions is also provided by the SSQBM tree structure.

In Chapter 4, MDSQ-SR-based MDC scheme was proposed that provides the quality scalability in each description. In MDSQR-SR, the base layer of each side description was successively refined to provide the quality scalability both in side and in central decoding. The base layer was designed by a well-known MDSQ method and different index assignment matrices were used to add different amounts of redundancy between the descriptions. The objective of the MDSQ-SR design was to improve the distortion for every refinement layer of a side description when individually decoded and for any combination of levels of refinement of the two refined side descriptions for joint decoding. The relationship between the design parameters, the refinement factor r of the side quantizer and the side quantizer spread factor g to meet the distortion constraints of the proposed MDSQ-SR was also derived in Chapter 4. It is proposed that the design conditions, $r > 1$, $r > g$ and $r^p \neq zg$, where z is a positive integer and p is the refinement levels, are jointly sufficient and necessary for satisfying the distortion constraints and verified through simulation results. The performance of the MDSQ-SR for various base layer options was evaluated and compared with EMDSQ and single description coding. The unbalanced and balanced joint decoding in terms of quality layers of the MDSQ-SR-based MDC scheme has shown the average PSNR improvement of 1.66 dB and 1.35 dB with respect to the joint decoding of the EMDSQ-based MDC scheme. Different amount of redundancy between the descriptions can be added by using different index assignment matrices at the base layer of MDSQ-SR.

In Chapter 5, both the resolution and quality scalability in multiple description image coding was achieved by integrating MUDC and MDSQ-SR respectively.

Different spatial resolution descriptions are generated from several MDSQs in such a way that the distortion constraints defined for MUDC are satisfied. Similarly, quality scalability in each spatial resolution description is achieved by successive refinement of the side quantizers. It was evident from the rate distortion performance that the joint decoding from different resolution descriptions is better than the joint decoding of the same spatial resolution descriptions. The joint decoding from quarter and full resolution descriptions has shown the PSNR improvement of 2.65 dB with respect to the joint decoding from full resolution descriptions. Similarly, the joint decoding from half and full resolution descriptions has shown the PSNR improvement of 1.6 dB with respect to the joint decoding from full resolution descriptions. However, the decoded quality of the joint decoding from different resolution description has shown rapid decay in quality than the joint decoding of similar resolution description under packet erasure channel. The side decoding quality and overall robustness was improved by using the concept of side information in MDSQ-based MDC scheme. It was evident from simulation results that the side decoding quality can be the same as central decoding quality. Similarly, by using the side information in MDC has shown 0.2-0.7 dB better performance under packet loss environment than the simple MDSQ-based MDC scheme.

In Chapter 6, a new MCTF-based framework for SMDVC using MDSQ-SR was presented. The MDSQ-SR is used to provide the quality scalability in each description, while the temporal decomposition facilitates to decode the video at different frame rate. The rate allocation scheme for each description is addressed by selecting different spatio-temporal subbands for each description in such a way that the joint decoding gives the minimum distortion. The joint decoding of the MDSQ-SR-based SMDVC has shown the average PSNR improvement of 0.3 dB with respect to the joint decoding of the EMDSQ-based MDVC. The problem of inter description redundancy control in SMDVC was also considered in this chapter. The amount of redundancy in texture is reduced by high frequency frame splitting among different descriptions. The joint decoding with temporal frame splitting has shown 0.6 dB better performance that the joint decoding by considering all frames in each description. The motion vector redundancy is reduced by generating multiple motion vector streams using MDSQ. The redundancy between the descriptions can be reduced by temporal frame splitting in SMDVC but with the sacrifice of 0.2-0.6 dBs in robustness. However, the multiple streams of the motion vectors generated from MDSQ not only reduced the data rate of each description but also showed better robustness when compared with frame splitting method. Finally, the integration of the MDSQ-SR and MUDC schemes in MCTF-based video coding framework was evaluated for P2P video streaming. By integrating these two schemes, different scalable descriptions can be extracted for each peer according to the peer uplink bandwidth. The proposed SMDVC scheme is capable of balanced and unbalanced joint decoding from the same and different MDSQ-SR at the receiving peer. It was evident from simulation results that the proposed SMDVC results in high robustness to packet losses as well as peer losses in P2P video streaming when compared with state-of-the-art MDSQ-based MDVC and single scalable description coding schemes.

Finally, it can be concluded that the proposed MUDC and MDSQ-SR based framework of SMDVC in this research is capable of generating any number of scalable descriptions that can be decoded at any quality, resolution and frame rate. The proposed framework facilitates the user to truncate the scalable descriptions in balanced and unbalanced manner by considering same and different quality layers, spatial resolution and frame rate in each description. The joint decoding of the balanced and unbalanced descriptions always shows the improvement in the quality, resolution and frame rate. The comparison of the proposed framework for achieving different scalabilities with other methods available in the literature has shown better performance in terms of rate-distortion and robustness. From the complexity point of view, the encoding time of creating multiple scalable descriptions by the proposed method is increased with respect to the single scalable description coding. The increase in the complexity is due to the MDSQ-SR block, which increase the encoding time by 1.31 sec/frame when compared with the single scalable description coding. On the other hand, the encoding time of the MDSQ-SR based SMDVC framework is reduced by 4.04 sec/frame with respect to the EMDSQ based scheme. The proposed MDSQ-SR based SMDVC framework not only performs well in terms of rate-distortion and robustness but also reduces the computational complexity by 83% when compared with the state-of-the-art EMDSQ based SMDVC scheme.

7.1 Key Contributions

The research presented in this thesis produced the following novel contributions in the field of multiple description image and video coding:

- Proposed and verified the distortion constraints and design conditions for generating any number of descriptions using more than one MDSQs that are capable of joint decoding in balanced and unbalanced manner.
- Proposed and verified the distortion constraints and design conditions for novel MDSQ-SR scheme, which is capable of generating quality scalable descriptions.
- Proposed a new joint decoding criteria of different spatial resolution descriptions with the objective to improve the spatial resolution and decoding quality.
- Used the side information in the base layer of MDSQ-SR to improve the side decoding quality and overall robustness of the MDC system.
- Proposed a novel MCTF based scalable multiple description video coding framework that is capable of generating any number of quality, resolution and frame rate scalable descriptions.
- Performance evaluation and comparison of the proposed scalable multiple description video coding scheme for peer to peer video streaming.

7.2 Future Work

The research presented in this thesis can be extended to pursue further research in multiple description image and video coding domain. Following are the list of future directions:

Channel Adaptive Multiple Description Video Coding : The rate and redundancy allocation criteria in the proposed SMDVC is based on minimizing the overall distortion of joint decoding without any prior knowledge of channel conditions. Therefore, the redundancy allocation in each description is fixed for any channel conditions. The robustness of the MDC scheme can be enhanced by considering adaptive rate and redundancy allocation according to the channel conditions. For this purpose, the rate allocation problem can be formulated by minimizing the expected distortion for any target bit rate and channel conditions. By considering the channel conditions in rate allocation of each description, more coding gain, efficient redundancy control mechanism and more robustness against channel conditions can be achieved.

ROI based MDC : In images and video there are certain parts that are of greater importance than others. This feature of images and video can be exploited to define certain ROIs in image and video to be encoded more securely and in better quality. Better rate and redundancy can be allocated in each description by using ROI-based coding. By using ROI, more number of descriptions and quality layers can be used for the interested regions in image and video. Similarly, low redundancies in terms of number of descriptions and quality layers can be used for uninterested regions resulting in little excess to the joint rate. By allocating different redundancies in different regions of image and video makes it applicable for low data rate applications.

Scalable and Reliable Transport of 3D Video over Packet Networks : 3D video is considered as a next evolution in multimedia technologies. There are different ways to represent the 3D video. But view plus depth based representation is usually used for efficient transmission of 3D video. In view plus depth representation, a single view and its associated depth map are transmitted and then rendered at the decoder side to generate both the views. However, in order to reproduce 3D video properly the depth map information needs to be accurately transmitted. Scalable multiple description video coding can be used as a solution for reliable transmission of depth map information over packet networks. The SMDVC framework proposed in this thesis is for 2D video, which can be extended to 3D video that provides the scalability and reliability in the transport of 3D video over best effort networks. In order to incorporate the scalability in 3D careful consideration is required for view scalability in each description.

Appendix A

Experimental Setup

Figure A.1 shows the block diagram of the test model used in this thesis for simulation for $N = 2$ descriptions. MDC encoder generates two scalable descriptions of the input image or video. The packetizer block is responsible for making packets from the scalable descriptions. Packets from each scalable description are then transmitted through two different channels. The details of the MDC encoder for image and video, packetizer and channels are discussed in detail in next sections.

A.1 MDC Encoder

Figure A.2 and Figure A.3 shows the block diagram of the MDC encoder used in this thesis for simulation for images and video respectively. In Figure A.2, the input image is decorrelated by using the 5/3 DWT. The number of decomposition levels of the DWT used in simulation is 5. The transformed coefficients are then quantized by using MDSQ-SR. The parameters of the MDSQ-SR *i.e.,* (a, g, r, and P) are used according to the conditions proposed in Chapter 3 and Chapter 4. The values of the parameters used for simulation are $a = 3$, $g = 2$, $r = 3$ and $P = 4$. The descriptions generated by the MDSQ-SR are then entropy coded by using Huffman coding. All the blocks used in the simulations are developed in Matlab by the author.

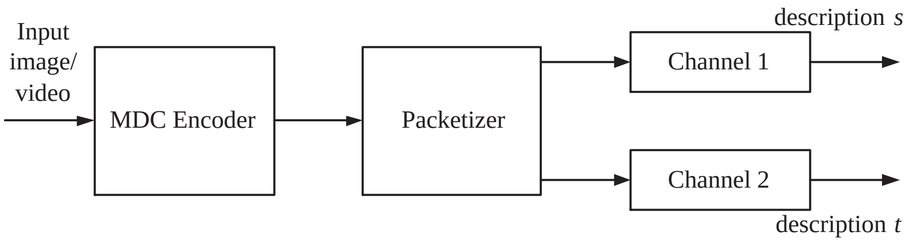


Figure A.1: Block diagram of test model.

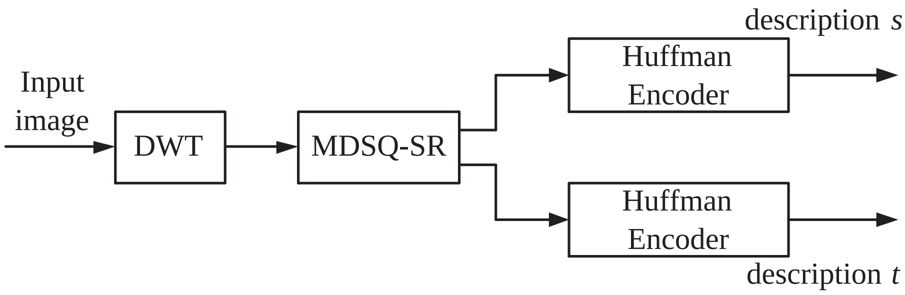


Figure A.2: Block diagram of MDC encoder for images.

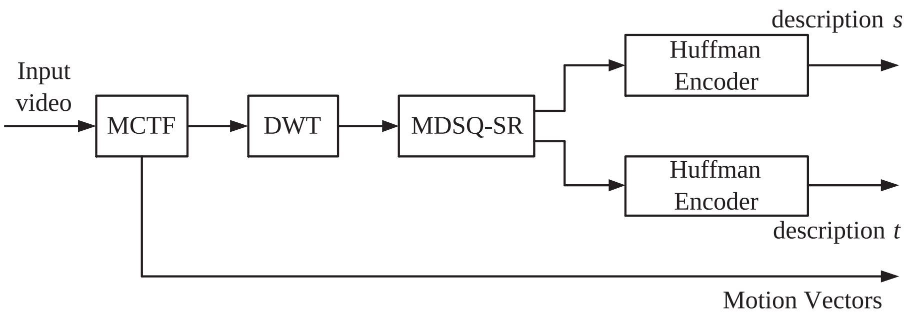


Figure A.3: Block diagram of MDC encoder for video.

Similarly, for multiple description video coding, input video sequence is temporally decomposed into low and high frequency frames and set of motion vectors by using MCTF. The GOP size of 16 and the temporal decomposition levels of 4 are used in simulations. The temporally decomposed frames are then decorrelated by using 5/3 DWT. The spatio-temporal coefficients are then quantized by using MDSQ-SR. The parameters of the DWT and MDSQ-SR are the same as used in MDC encoder of images. The quantized spatio-temporal coefficients are then entropy coded by using Huffman coding. The MCTF block used for simulation is developed by [2] and other blocks are developed by the author.

A.2 Packetizer and Channel Model

The main purpose of the packetizer block is to packetized each description generated from the MDSQ-SR. In this thesis, wavelet tree based packetization is used, in which single wavelet tree is considered as a one packet. For video sequence, spatio-temporal wavelet tree is considered as a one packet. Furthermore, the refinement information of each wavelet tree is placed in the same packet to minimize the erasure effect. The reason of placing the refinement information in the same packet is that if the base layer packet is lost during transmission then there is no advantage of receiving the refinement information.

After packetization, the packets are transmitted through two different channels. In this thesis, packet erasure channels with random packet loss patterns for any percentage of packet loss are used for simulation. Let M be the total number of packets for each description and l be the number of lost packets on any random packet channel. Then there is a total of ${}^{M}C_{l}$ number of combinations to loose l packets from the total of M packets. The Matlab program developed by the author generates all packet loss patterns for any particular percentage of packet loss. The effect of packet loss on image quality is calculated for every packet loss pattern at any particular percentage of packet loss. However, for the video, the effect of packet loss on video quality is calculated only for 20-100 different random packet loss patterns selected randomly from all the combinations of packet loss patterns generated by the Matlab program.

A.3 Test Image/ Video Sequence Set and Performance Evaluation Criteria

The test image set used for simulations in this thesis consists of five images, namely, *Gold Hill*, *Barbara (on chair)*, *Barbara (on floor)*, *Blackboard* and *Boats*. All the images used are gray scale and of dimension 576×704. The test video sequences set used for simulation consist of four sequences *i.e., Mobile*, *Foreman*, *Bus* and *Harbour*. All the sequences used are coloured sequences and of CIF format *i.e.,* (352×288, 30 frames/sec). Figure A.4 and Figure A.5 shows the test

Gold Hill

Barbara (on chair)

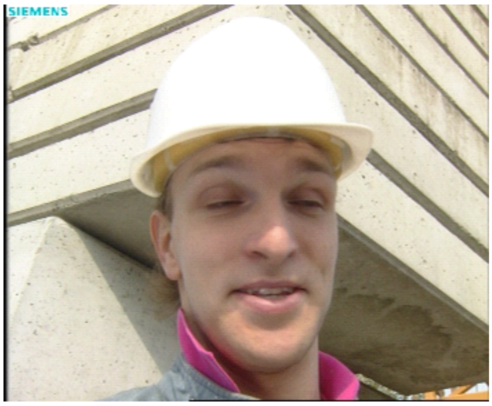
Barbara (on floor)

Blackboard

Boats

Figure A.4: Image data set.

Mobile



Foreman

Bus

Harbour

Figure A.5: Video sequence data set.

image and video sequence set respectively. The images selected for the test data set are based on different texture information. Similarly, the video sequences selected for the test data set are based on different texture and motion information. In this thesis multiple scalable descriptions of video are generated only for the texture information using MDSQ-SR and the same motion vector information is used in each description. Also, in the simulation it is assumed that the motion vector information is available at the decoder without any losses which means the effect on the decoded video quality is due to the texture information. Therefore, the video sequences used in this thesis have low or medium motion.

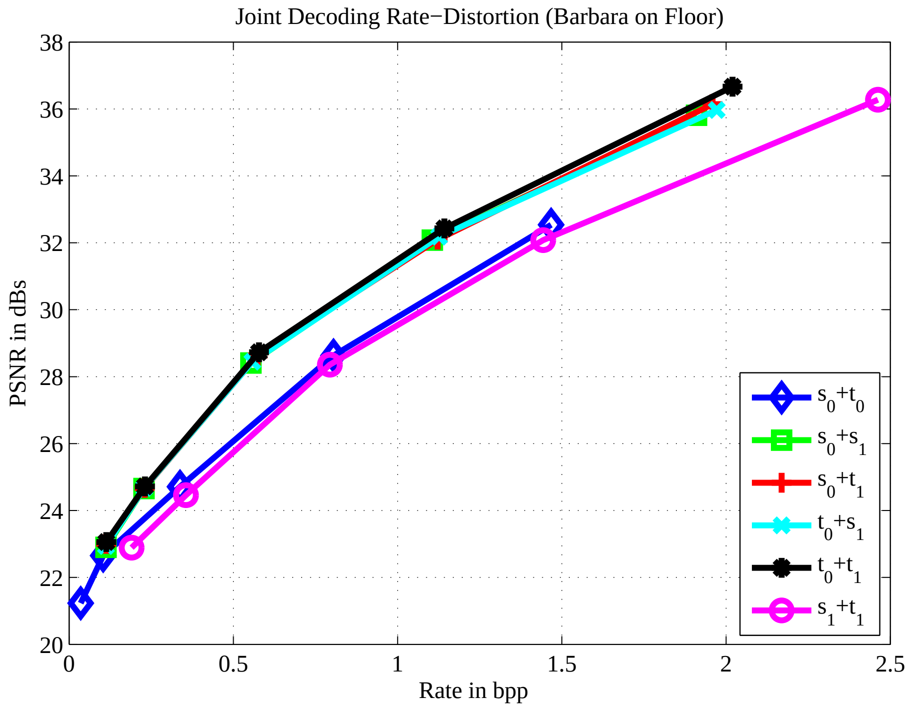
The schemes presented in this thesis are evaluated under lossless and lossy channel conditions. For lossless channel conditions, rate distortion performance of the proposed schemes for images and video are presented in the form of rate distortion curves. For lossy channel conditions, packet erasure channels are considered to evaluate the effect of packet loss on the decoded quality of image or video and results are presented in the form of loss distortion curves. For rate distortion curves rate is calculated in terms of bits per pixel (bpp) and bits per second (bps) for images and video respectively. For loss distortion curves loss is considered in terms of percentage packet loss in each description. The most common metric used for measuring quality of image and video is peak signal to noise ratio (PSNR) due to its simplicity. Therefore, for rate distortion and loss distortion curves, distortion is calculated in terms of PSNR by using the Eq. (2.2). The PSNR values for low and high image and video qualities lies between 30 dB and 40 dB respectively. Therefore, small improvement in PSNR at a particular rate is considered as significant improvement.

In PSNR calculation, average error between the original and compressed frame is considered and no information regarding human visual system is used. Video quality metric (VQM) and structural similarity measure (SSIM) are the metrics for measuring video quality that considers human visual system for extracting structural information from a scene. In this thesis, the VQM and SSIM results of the video sequences are also presented in Chapter 6. The VQM and SSIM results are obtained by using the MSU video quality measurement tool.

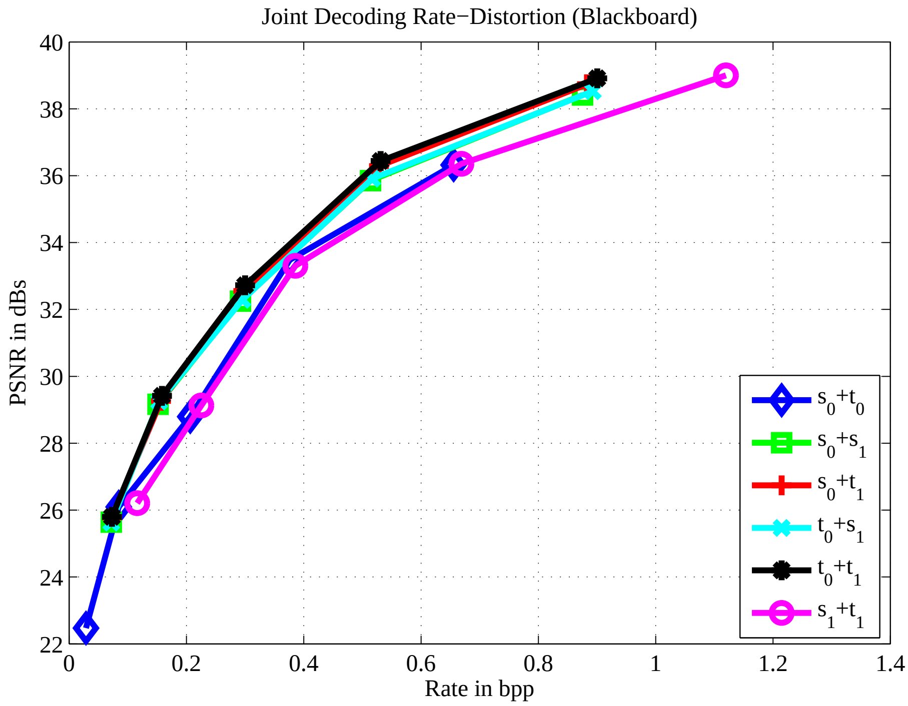
Appendix B

Additional Results

In this appendix, some additional rate distortion results of multichannel unbalanced description coding and multiple description scalar quantizer with successive refinement are presented for *Barbara (on floor)*, *Blackboard* and *Boats* images.

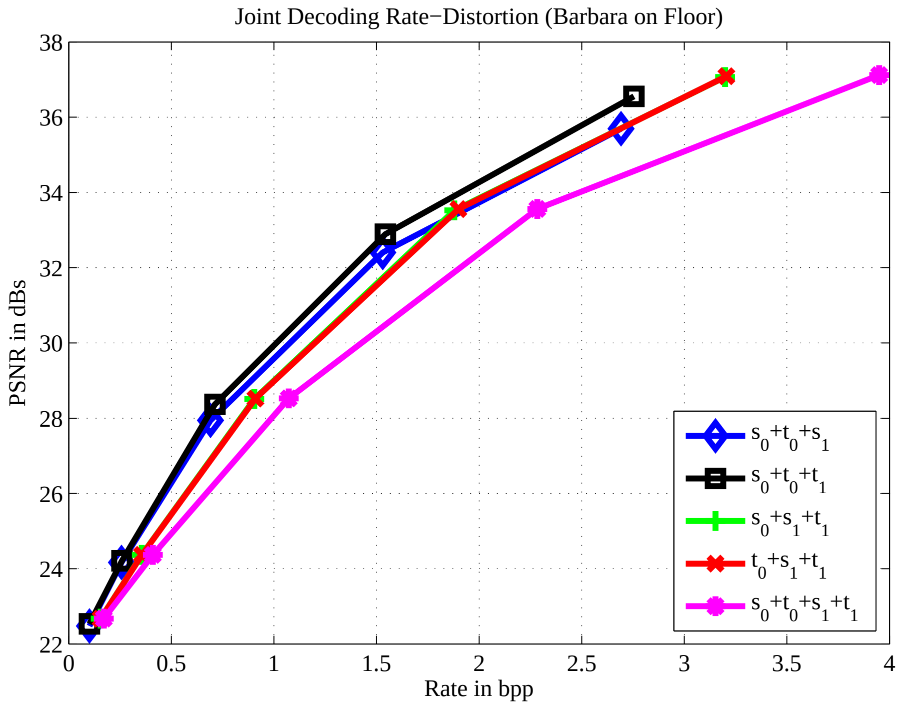


(a)

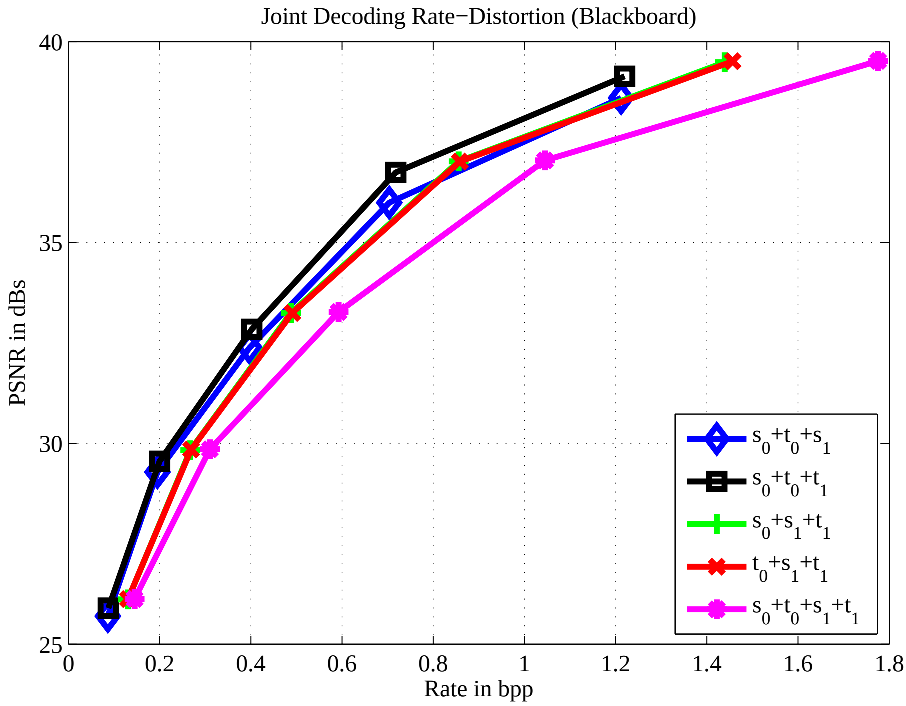


(b)

Figure B.1: Joint decoding rate-distortion curves for $N = 2$ descriptions from $J = 2$ MDSQs for (a) *Barbara (on floor)*(b) *Blackboard* images.

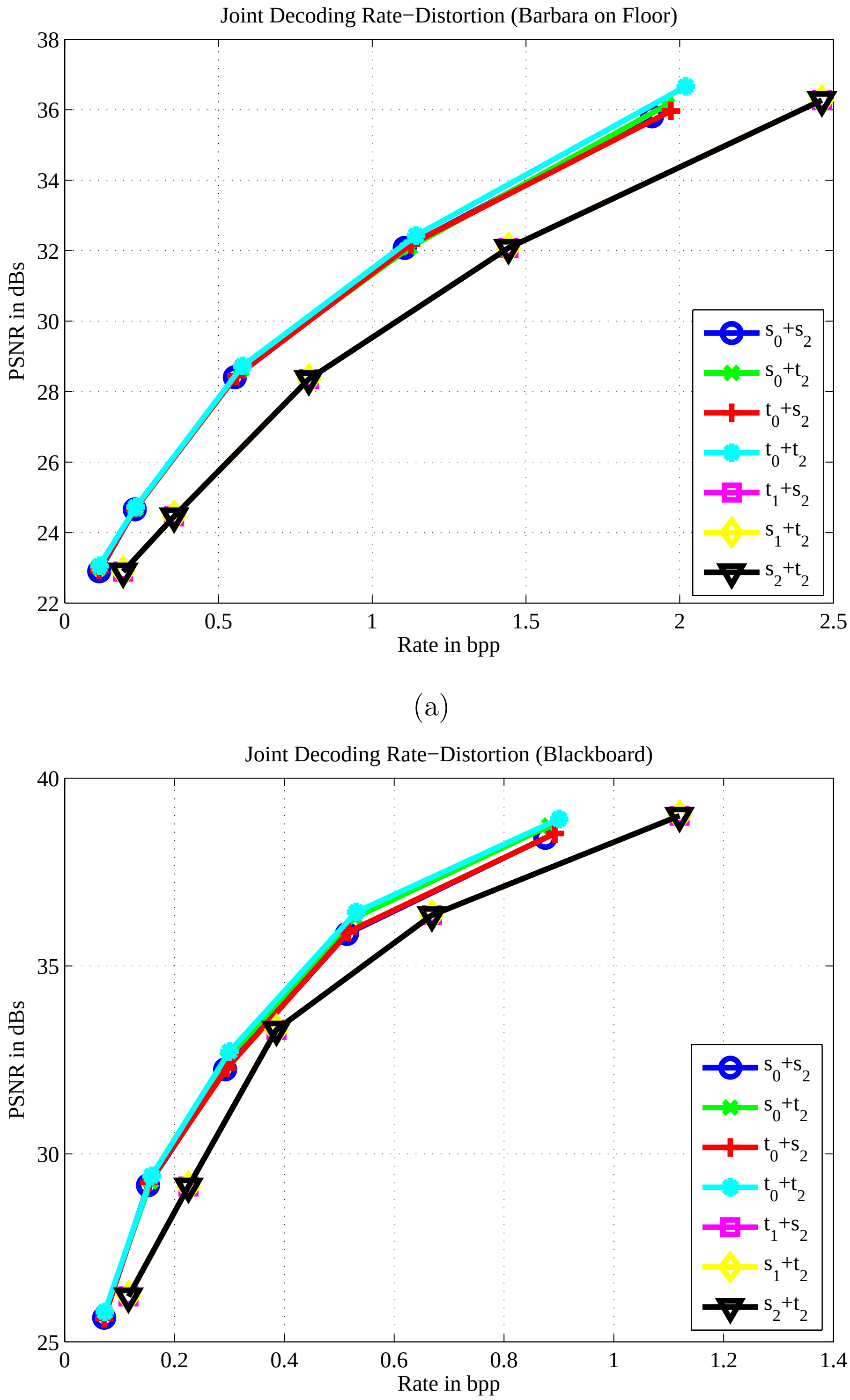


(a)



(b)

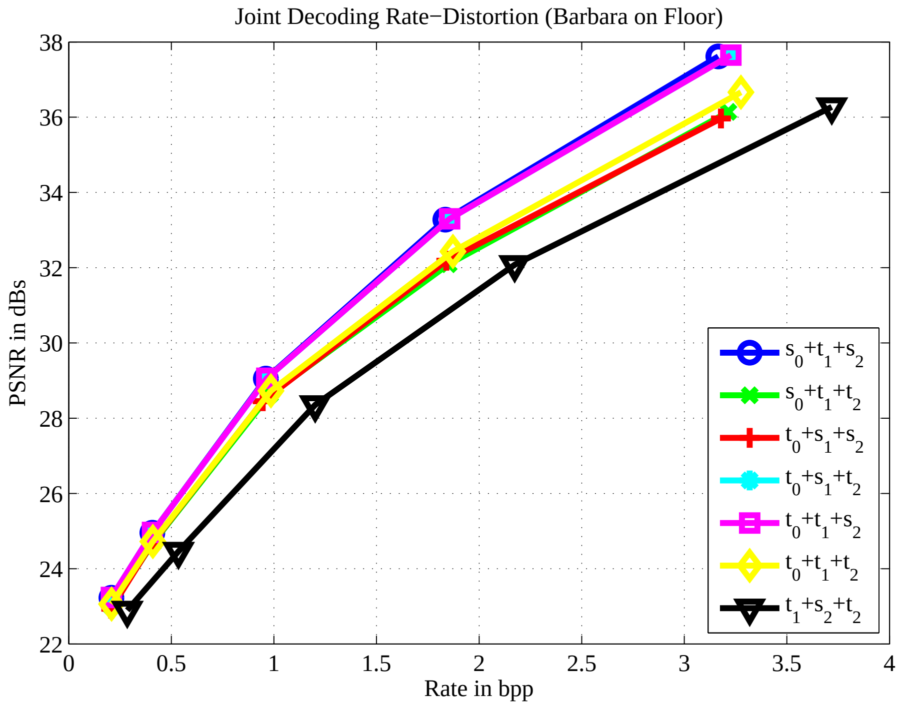
Figure B.2: Joint decoding rate-distortion curves for $N = 3, 4$ descriptions from $J = 2$ MDSQs for (a) *Barbara (on floor)*(b) *Blackboard* images.



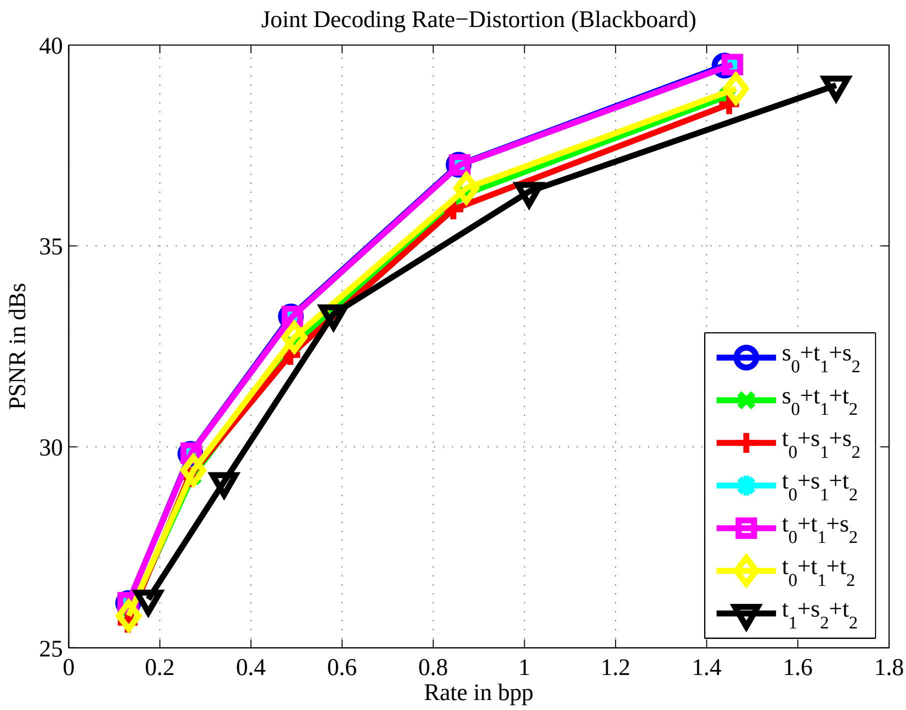
(a)

(b)

Figure B.3: Joint decoding rate-distortion curves for $N = 2$ descriptions from $J = 3$ MDSQs for (a) *Barbara (on floor)* (b) *Blackboard* images.

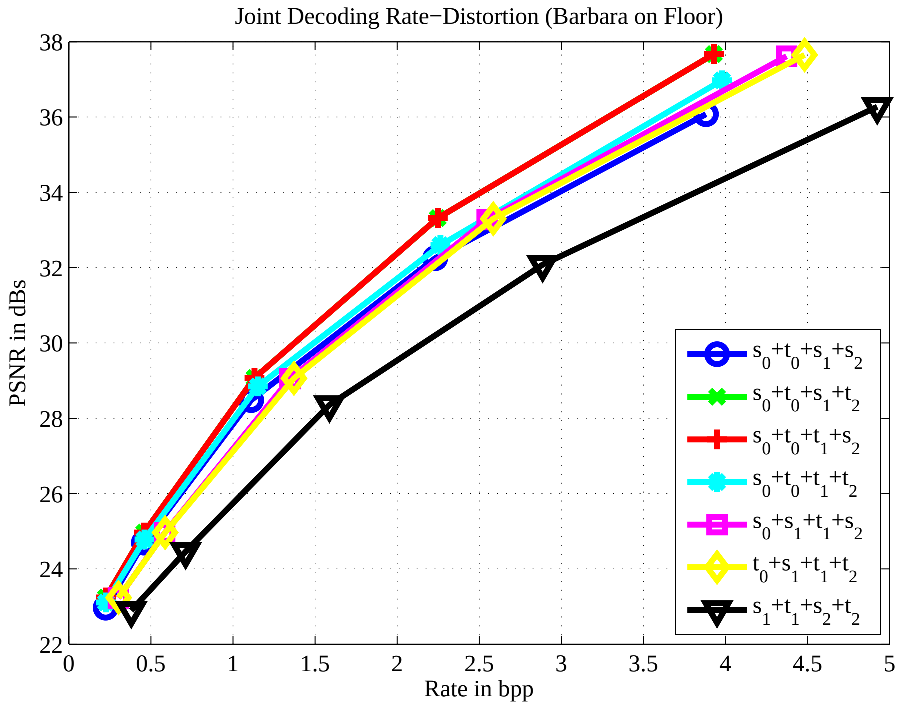


(a)

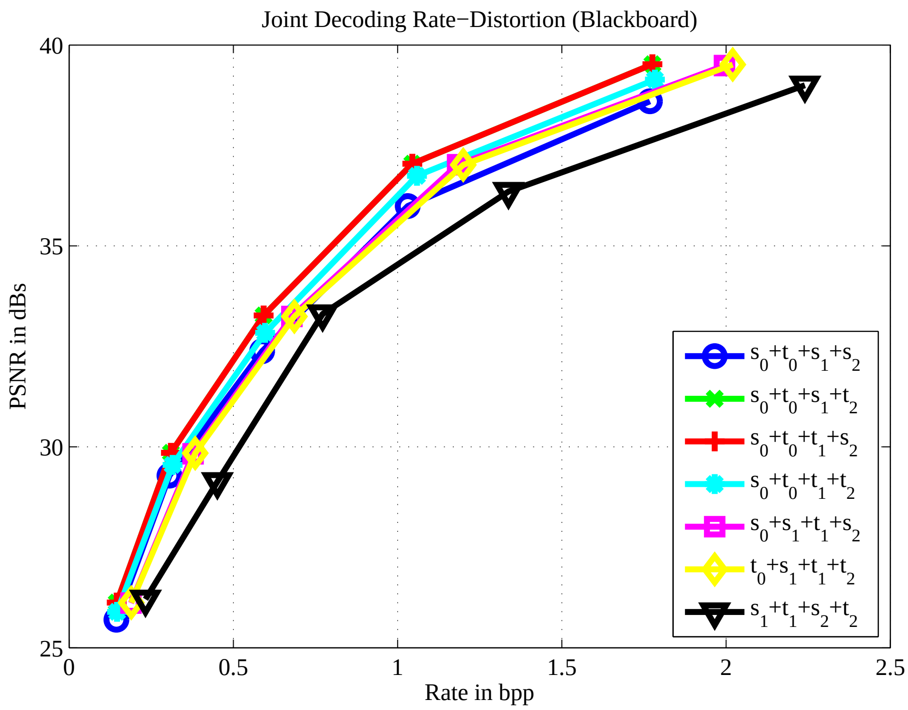


(b)

Figure B.4: Joint decoding rate-distortion curves for $N = 3$ descriptions from $J = 3$ MDSQs for (a) *Barbara (on floor)* (b) *Blackboard* images.

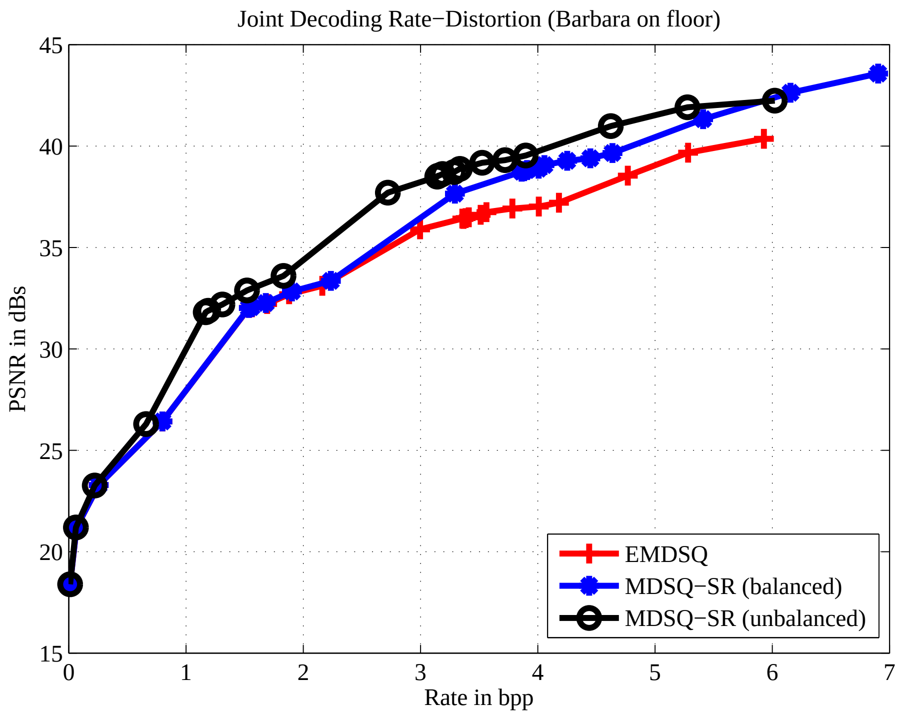


(a)

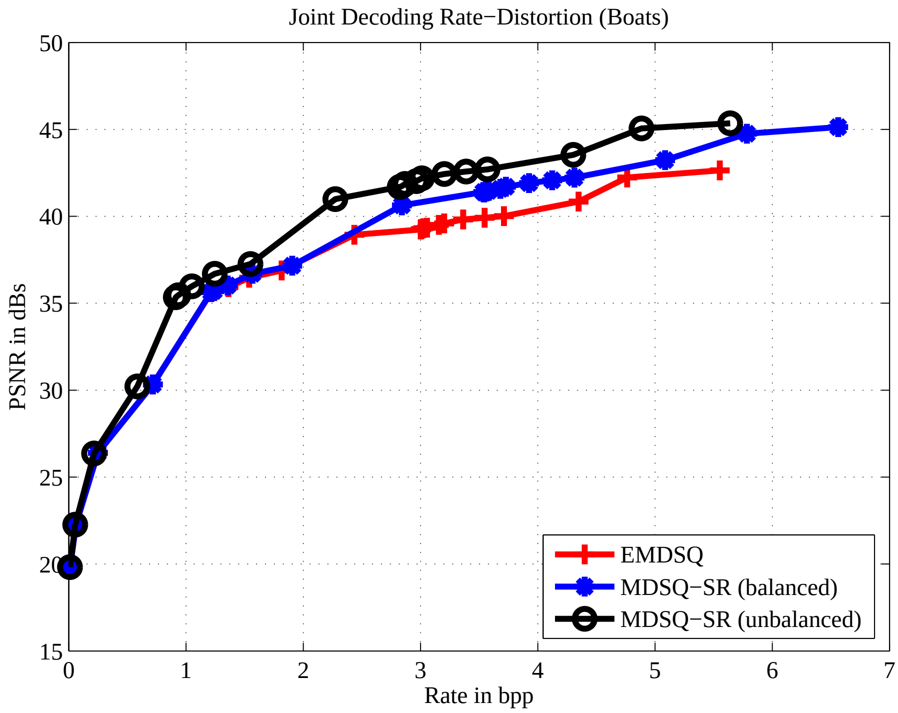


(b)

Figure B.5: Joint decoding rate-distortion curves for $N = 4$ descriptions from $J = 3$ MDSQs for (a) *Barbara (on floor)* (b) *Blackboard* images.

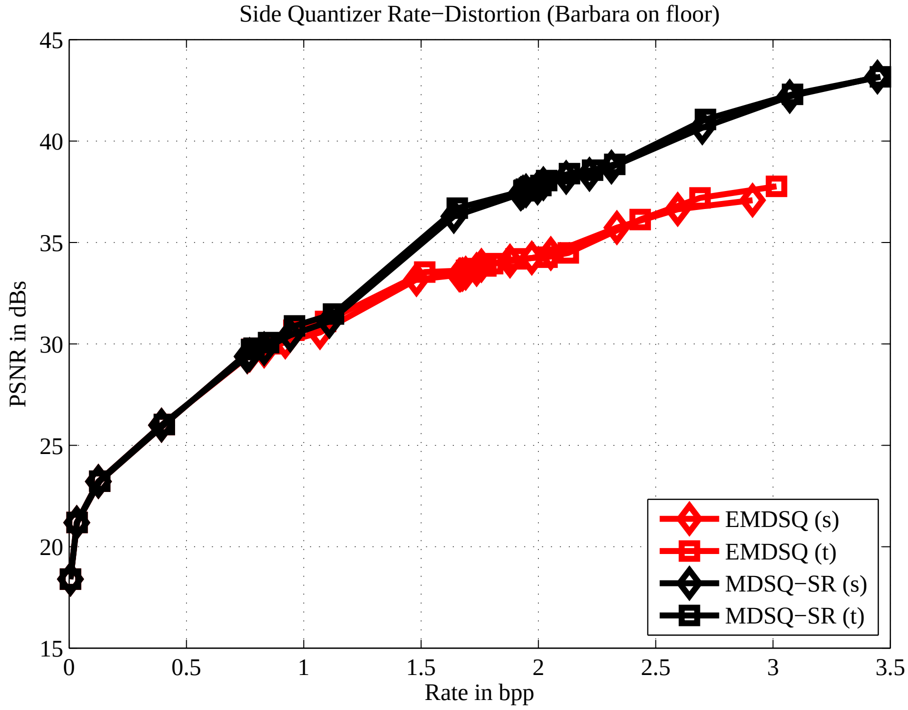


(a)

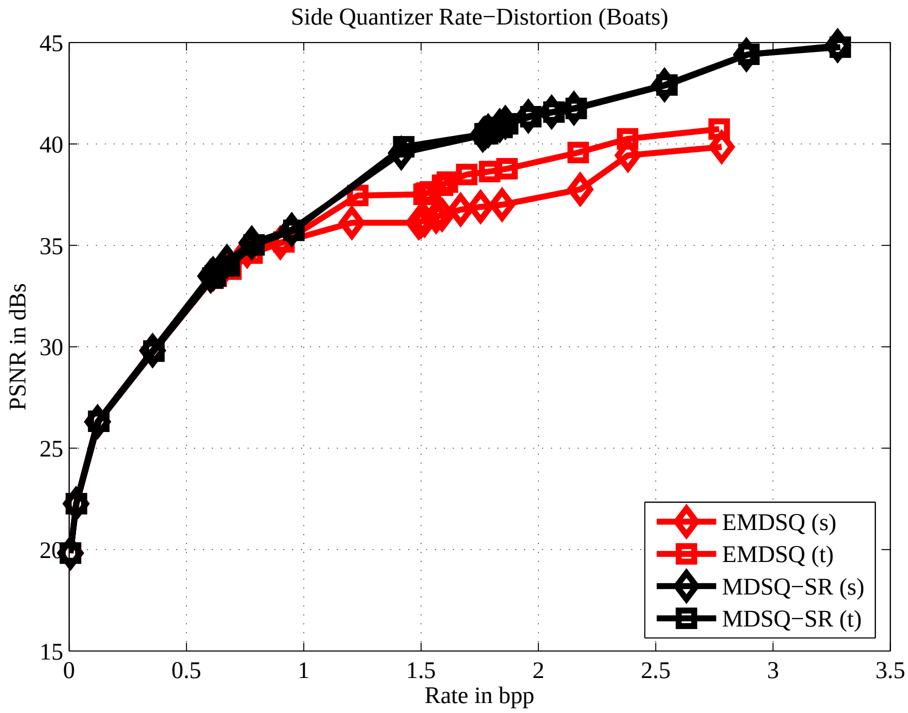


(b)

Figure B.6: Joint decoding comparison of the proposed MDSQ-SR and EMDSQ schemes for (a) *Barbara (on floor)* (b) *Boats* images.

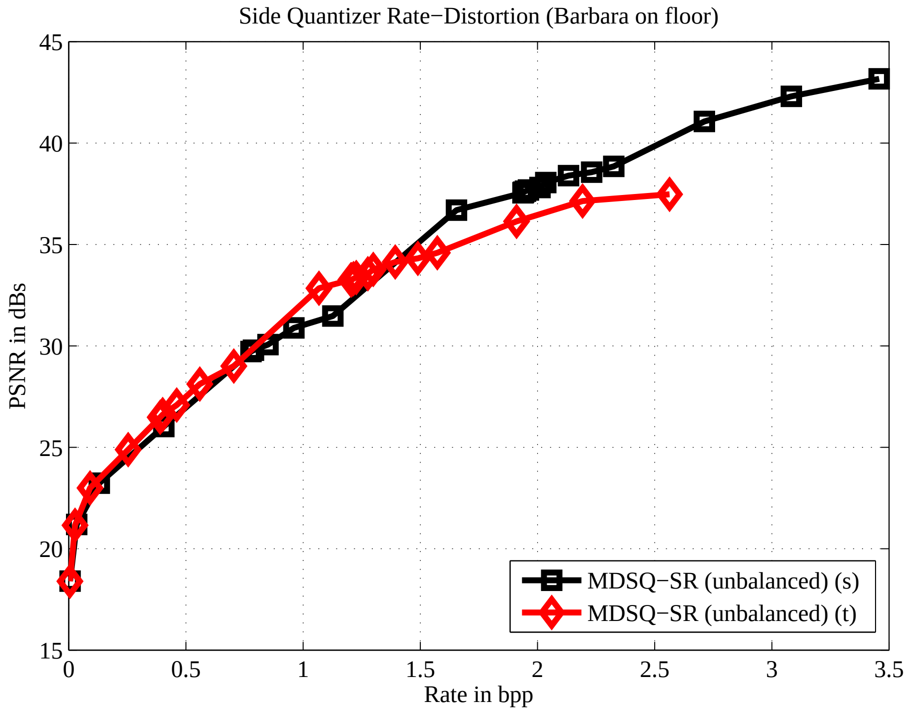


(a)

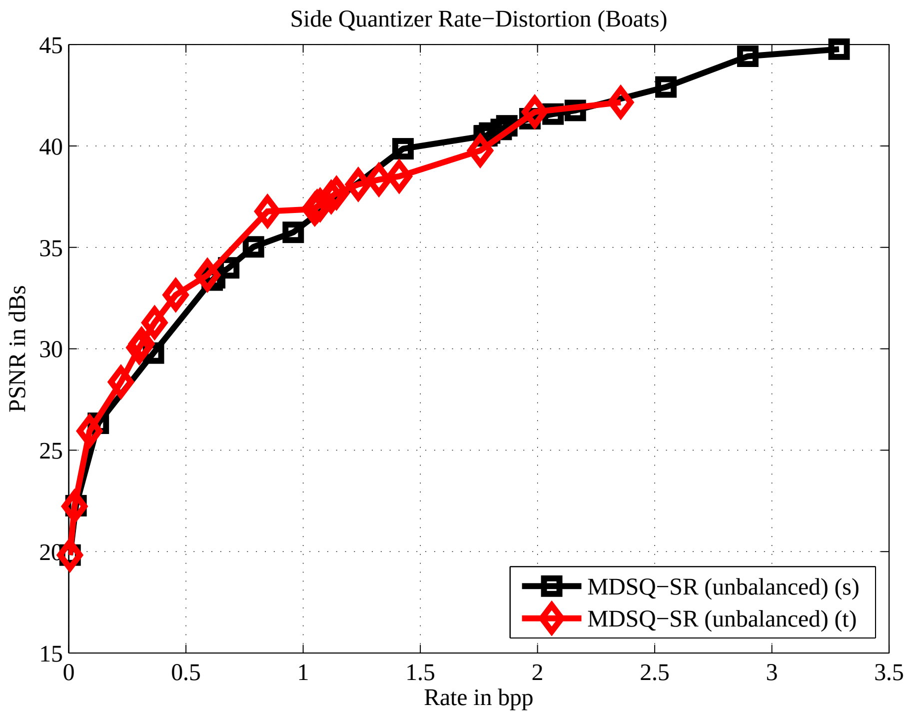


(b)

Figure B.7: Balanced side description decoding comparison of the proposed MDSQ-SR and EMDSQ schemes for (a) *Barbara (on floor)* (b) *Boats* images.



(a)



(b)

Figure B.8: Unbalanced side description decoding of the proposed MDSQ-SR schemes for (a) *Barbara (on floor)* (b) *Boats* images.

References

[1] N. Adami, A. Signoroni, and R. Leonardi, "State-of-the-art trends in scalable video compression with wavelet-based approaches," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1238–1255, 2007.

[2] S. T. Hsiang and J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank," *Signal Processing: Image Communications*, vol. 16, no. 8, pp. 705–724, 2001.

[3] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magzine*, pp. 74–93, 2001.

[4] V. K. Goyal and J. Kovacevic, "Generalized multiple description coding with correlating transform," *IEEE Trans. on Information Theory*, vol. 47, no. 6, pp. 2199–2224, 2001.

[5] S. D. Servetto, K. Ramchandran, V. A. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Trans. on Image Processing*, vol. 9, no. 5, pp. 813–826, 2000.

[6] T. Tillo and A. Olmo, "Data-dependent pre- and postprocessing multiple description coding of images," *IEEE Trans. on Image Processing*, vol. 16, no. 5, pp. 1269–1280, 2007.

[7] N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro, "Multiple description video coding for scalable and robust transmission over IP," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 15, no. 3, pp. 321–334, 2005.

[8] L. P. Kondi, "A rate-distortion optimal hybrid scalable/multiple description video codec," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 15, no. 7, pp. 921–927, 2005.

[9] S. Adedoyin, W. A. C. Fernando, H. A. Karim, C. T. E. R. Hewage, and A. M. Kondoz, "Scalable multiple description coding with side information using motion interpolation," *IEEE Trans. on Consumer Electronics*, vol. 54, no. 4, pp. 2045–2052, 2008.

[10] I. V. Bajic and J. W. Woods, "Domain-based multiple description coding of images and video," *IEEE Trans. on Image Processing*, vol. 12, no. 10, pp. 1211–1225, 2003.

[11] M. van der Schaar and D. Turaga, "Multiple description scalable coding using wavelet-based motion compensated temporal filtering," in *Proc. IEEE Int'l Conf. on Image Processing*, 2003, pp. 489–492.

[12] F. Verdicchio, A. Munteanu, A. I. Gavrilescu, J. Cornelis, and P. Schelkens, "Embedded multiple descripiton coding of video," *IEEE Trans. on Image Processing*, vol. 15, no. 10, pp. 3114–3130, 2006.

[13] H. A. Karim, C. Kodikara, S. Worrall, A. H. Sadka, and A. M. Kondoz, "Multiple description video coding using frame interpolation," in *Special VISNET Session at Polish National Conference*, 2005, pp. 53–59.

[14] H. A. Karim, C. T. E. R. Hewage, S. Worrall, and A. M. Kondoz, "Scalable multiple description video coding for stereoscopic 3D," *IEEE Trans. on Consumer Electronics*, vol. 54, no. 2, pp. 745–752, 2008.

[15] H. Wang and A. Ortega, "Robust video communication by combining scalability and multiple description coding techniques," in *Proc. of SPIE*, vol. 5022, 2003, pp. 111–124.

[16] H. Mansour, P. Nasiopoulos, and V. Leung, "An efficient multiple desccription coding scheme for the scalable extension of H.264/AVC (SVC)," in *Symposium on Signal Processing and Information Technology*, 2006, pp. 519–523.

[17] T. B. Abanoz and A. M. Teklap, "SVC-based scalable multiple descripiton video coding and optimization of encoding configuration," *Signal Processing: Image Communications*, vol. 24, pp. 691–701, 1998.

[18] M. L. Hilton, B. D. Jawerth, and A. Sengupta, "Compressing still and moving images with wavelets," *Multimedia Systems*, vol. 2, no. 5, pp. 218–227, 1994.

[19] R. M. Gray, "Vector quantization," *IEEE ASSP Magzine*, vol. IT-1, pp. 4–29, 1984.

[20] D. A. Huffman, "A method for the construction of minimum-redundancy codes," in *Proc. of IRE*, vol. 40, 1952, pp. 1098–1101.

[21] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Communications of ACM*, vol. 30, no. 6, pp. 520–540, 1987.

[22] A. N. Netravali and J. D. Robbins, "Motion compensated television coding: Part I," *Bell Syst. Tech. J.*, vol. 58, no. 3, pp. 631–670, 1979.

[23] J. R. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. on Comm.*, vol. COM-29, pp. 1799–1808, 1981.

[24] G. K. Wallace, "The JPEG still picture compression standard," *Communications of ACM*, vol. 34, pp. 31–44, 1991.

[25] C. A. Christopoulos, A. N. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: An overview," *IEEE Trans. on Consumer Electronics*, vol. 46, pp. 1103–1127, 2000.

[26] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG2000 still image compression standard," *IEEE Signal Processing Magzine*, pp. 36–58, 2001.

[27] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[28] G. J. Sullivan and T. Wiegand, "Video compression - from concepts to the H.264/AVC standard," *Proc. of the IEEE*, vol. 93, no. 1, pp. 18–31, 2005.

[29] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.

[30] N. Sprljan, M. Mrak, G. Abhayaratne, and E. Izquierdo, “A scalable coding framework for efficient video adaptation,” in *Proc. 6th Int’l Workshop on Image Analysis for Multimedia Interactive Services*, 2005.

[31] J. M. Shapiro, “Embedded image coding using zerotrees of wavelet coefficients,” *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.

[32] A. Said and W. A. Pearlman, “A new fast and efficient image codec based on set partitioninig in hierarchical trees,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, 1996.

[33] D. Taubman, “High performance scalable image compression with EBCOT,” *IEEE Trans. on Image Processing*, vol. 9, no. 7, pp. 1158–1170, 2000.

[34] E. Akyol, A. M. Tekalp, and M. R. Civanlar, “Motion compensated temporal filtering witnin the H.264/AVC standard,” in *Proc. IEEE Int’l Conf. on Image Processing*, vol. 4, 2004, pp. 2291–2294.

[35] S. J. Choi and J. W. Woods, “Motion compensated 3-D subband coding of video,” *IEEE Trans. on Image Processing*, vol. 8, no. 2, pp. 155–167, 1999.

[36] B. Pesquet-Popescu and V. Bottreau, “Three-dimensional lifting scheme for motion-compensated video compression,” in *Proc. IEEE Int’l. Conf. on Acoustics, Speech, and Signal Processing*, 2001, pp. 1793–1796.

[37] L. Luo, J. Li, S. Li, Z. Zhuang, and Y. Q. Zhang, “Motion compensated lifting wavelet and its application in video coding,” in *Proc. IEEE Int’l. Conf. on Multimedia and Expo*, 2001, pp. 365–368.

[38] A. Secker and D. Taubman, “Motion-compensated highly scalable video compression using and adaptive 3d wavelet transform based on lifting,” in *Proc. IEEE Int’l Conf. on Image Processing*, vol. 2, 2001, pp. 1029–1032.

[39] M. F. Lopez, S. G. Rodriguez, J. P. Ortiz, J. M. Dana, Vicente, G. Ruiz, and I. Garcia, “Fully scalable video coding with packed stream,” in *Image and Video Communication and Processing*, vol. 5685(2), 2005, pp. 378–389.

[40] A. Secker and D. Taubman, “Lifting-based invertible motion adaptive transform framework for highly scalable video compression,” *IEEE Trans. on Image Processing*, vol. 12, no. 12, pp. 1530–1542, 2003.

[41] J. R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. on Image Processing*, vol. 3, no. 5, pp. 559–571, 1994.

[42] C. C. Lin, Y. T. Hwang, K. H. Tseng, and S. W. Chen, "Wavelet based lossless video compression using motion compensated temporal filtering," in *Proc. IEEE Workshop on Signal Processing Systems*, 2007, pp. 686–691.

[43] D. Taubman and A. Zakhor, "Multirate 3D subband coding of video," *IEEE Trans. on Image Processing*, vol. 3, no. 5, pp. 572–589, 1994.

[44] Y. Andreopoulos, M. van der Schaar, A. Munteanu, J. Barbarien, P. Schelkens, and J. Cornelis, "Complete-to-over-complete discrete wavelet transform for scalable video coding with MCTF," in *Proc. SPIE Visual Communication and Image Processing*, vol. 5150, 2003, pp. 719–731.

[45] S. S. Tsai and H.-M. Hang, "Motion information scalability for MC-EZBC," *Signal Processing: Image Communications*, vol. 19, pp. 675–684, 2001.

[46] Y. Wu and J. W. Woods, "Scalable motion vector coding based on CABAC for MC-EZBC," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 6, pp. 790–795, 2007.

[47] Q. Chen and T. Fisher, "Image coding using robust quantization for noisy digital transmission," *IEEE Trans. on Image Processing*, vol. 7, no. 4, pp. 496–505, 1998.

[48] P. Sherwood and K. Zeger, "Progressive image coding for noisy channels," *IEEE Signal Processing Letter*, vol. 4, no. 7, pp. 189–191, 1997.

[49] ——, "Error protection for progressive image transmission over memoryless and fading channels," *IEEE Trans. on Comm.*, vol. 46, no. 12, pp. 1555–1559, 1998.

[50] S. Lin, D. Costello, and M. Miller, "Automatic repeat request error control schemes," *IEEE Trans. on Comm.*, vol. 22, no. 12, pp. 5–17, 1984.

[51] O. Crave, C. Guillemot, and B. Pesquet-Popescu, "Multiple scalar quantization with side information: Practical scheme and iterative decoding," *EURASIP J. on Advance Signal Processing*, vol. 2009, p. 10, 2009.

[52] O. Crave, C. Guillemot, B. Pesquet-Popescu, and C. Tillier, "Distributed temporal multiple description coding for robust video transmission," *EURASIP J. on Wireless Communication and Networking*, vol. 2008, p. 13, 2008.

[53] A. I. Gavrilescu, F. Verdicchio, A. Munteanu, I. Moerman, J. Cornelis, and P. Schelkens, "Scalable multiple-description image coding based on embedded quantization," *EURASIP Jnl. on Image and Video Processing*, vol. 2007, p. 11, 2007.

[54] A. I. Gavrilescu, A. Munteanu, J. Cornelis, and P. Schelkens, "Generalization of embedded multiple descripiton scalar quantizers," *IEE Electronics Letters*, vol. 41, no. 2, pp. 63–65, 2005.

[55] A. A. E. Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. on Information Theory*, vol. IT-28, no. 6, pp. 851–857, 1982.

[56] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, 1980.

[57] R. Ahlswede, "The rate distortion region for multiple descriptions without excess rate," *IEEE Trans. on Information Theory*, vol. IT-31, pp. 721–726, 1985.

[58] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. on Information Theory*, vol. IT-33, pp. 502–521, 1987.

[59] ——, "Multiple description source coding with no excess marginal rate," *IEEE Trans. on Information Theory*, vol. 41, pp. 349–357, 1995.

[60] B. Yin, Y. Zhang, Y. Shi, and D. Kong, "An improved scheme for multiple description image coding with Quincunx segmentation," in *Proc. IET Workshop on Visual Information Engineering*, 2008, pp. 783–788.

[61] N. Zhang, Y. Lu, F. Wu, and B. Yin, "Directional lifting-based wavelet transform for multiple description image coding with quincunx segmentation," *Lecture Notes on Computer Science*, vol. 3768, pp. 629–640, 2005.

[62] N. Franchi, M. Fumagalli, and R. Lancini, "Flexible redundancy insertion in a polyphase down sampling multiple description image coding," in *Proc. IEEE Int'l. Conf. on Multimedia and Expo*, vol. 2, 2002, pp. 605–608.

[63] S. Shirani, M. Gallani, and F. Kossentini, "Multiple description image coding using pre- and post-processing," in *Proc. Int'l. Conf. on Information Technology: Coding and Computing*, 2001, pp. 35–39.

[64] Y. Wang, M. T. Orchard, V. Vaishampayan, and A. R. Reibmen, "Multiple description coding using pairwise correlating transform," *IEEE Trans. on Image Processing*, vol. 10, no. 3, pp. 351–366, 2001.

[65] V. A. Vaishampayan, "Applications of multiple description codes to image and video transmission over lossy networks," in *Proc. Int'l. Workshop on Packet Video*, 1996.

[66] A. C. Miguel, A. E. Mohr, and E. A. Riskin, "SPIHT for generalized multiple description coding," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 3, 1999, pp. 842–846.

[67] V. A. Vaishampayan, "Design of multiple description scalar quantizer," *IEEE Trans. on Information Theory*, vol. 39, no. 3, pp. 821–834, 1993.

[68] K. Khelil, R. E. Bekka, A. Djebari, and J. M. Rouvaen, "Multiple description wavelet-based image coding using correlating transforms," *Int'l. Jnl. of Electronics and Communication*, vol. 61, no. 6, pp. 411–417, 2007.

[69] Y. Wang, M. Orchard, and A. R. Reibman, "Multiple description image coding for noisy channels by pairing transform coefficients," in *Proc. IEEE Workshop on Multimedia Signal Processing*, 1997, pp. 419–424.

[70] V. K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description transform coding of images," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 1, 1998, pp. 674–678.

[71] D. Chung and Y. Wang, "Multiple description image coding based on lapped orthogonal transform," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 1, 1998, pp. 664–668.

[72] G. Sun, U. Samarawickrama, J. Liang, C. Tian, C. Tu, and T. D. Tran, "Multiple description coding with prediction compensation," *IEEE Trans. on Image Processing*, vol. 18, no. 5, pp. 1037–1047, 2009.

[73] J. Chen, C. Cai, and K. Ma, "Layered tree-based multiple description coding," in *Proc. IEEE Int'l. Symposium on Intelligent Signal Processing and Communication Systems*, vol. 1, 2006, pp. 947–950.

[74] D. Song, L. Cao, and C. W. Chen, "Robust multiple description image coding over wireless networks based on wavelet tree coding, error resilient entropy coding, and error concealment," *Jnl. of Visual Communication and Image Representation*, vol. 19, no. 5, pp. 311–319, 2008.

[75] M. Wu, D. Song, and C. W. Chen, "Multiple descripiton distributed image coding with side information for mobile wireless transmission," *Image and Video Communication and Processing*, vol. 5685(2), pp. 572–582, 2005.

[76] V. A. Vaishampayan and J. Domaszewicz, "Design of entropy constrained multiple description scalar quantizer," *IEEE Trans. on Information Theory*, vol. 40, no. 1, pp. 245–250, 1994.

[77] V. A. Vaishampayan and J. Batllo, "Asymptotic analysis of multiple description quantizers," *IEEE Trans. on Information Theory*, vol. 44, no. 1, pp. 278–284, 1998.

[78] V. A. Vaishampayan, "Application of multiple descripiton codes to image and video transmission over lossy networks," in *Proc. Int'l. Workshop on Packet Video*, 1996, pp. 55–60.

[79] C. Tian, S. Mohajer, and S. N. Diggavi, "Approximating the gaussian multiple description rate region under symmetric distortion constraints," *IEEE Trans. on Information Theory*, vol. 55, no. 8, pp. 3869–3891, 2009.

[80] T. Guionnet, C. Guillemot, and S. Pateux, "Embedded multiple description coding for progressive image transmission over unreliable channels," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 1, 2001, pp. 94–97.

[81] A. I. Gavrilescu, A. Munteanu, P. Schelkens, and J. Cornelis, "Embedded multiple description scalar quantizers for progressive image transmission," in *Proc. IEEE Int'l. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, 2003, pp. 736–739.

[82] ——, "Embedded multiple description scalar quantizers," *IEE Electronics Letters*, vol. 39, no. 13, pp. 979–980, 2003.

[83] A. I. Gavrilescu, A. Munteanu, J. Cornelis, and P. Schelkens, "A new family of embedded multiple description scalar quantizers," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 1, 2004, pp. 159–162.

[84] M. Orchard, Y. Wang, V. Vaishampayan, and A. Reibman, “Redundancy rate distortion analysis of multiple description image coding using pairwise correlating transforms,” in *Proc. IEEE Int’l Conf. on Image Processing*, 1997, pp. 608–611.

[85] Y. Wang, M. Orchard, and A. Reibman, “Optimal pairwise correlating transform for multiple description coding,” in *Proc. IEEE Int’l Conf. on Image Processing*, vol. 1, 1998, pp. 679–683.

[86] Y. Wang and S. Lin, “Error-resilient coding using multiple description motion compensation,” in *Proc. IEEE Workshop on Multimedia Signal Processing (MMSP)*, 2001, pp. 441–446.

[87] P. Sagetong and A. Ortega, “Optimal bit allocation for channel-adaptive multiple description coding,” in *Proc. SPIE Visual Communication and Image Processing*, 2000, pp. 53–63.

[88] T. Tillo and G. Olmo, “A novel multiple description coding scheme compatible with the jpeg2000 decoder,” *IEEE Signal Processing Letter*, vol. 11, no. 11, pp. 908–911, 2004.

[89] Y. Wang, A. R. Reibmen, and S. Lin, “Multiple description coding for video delivery,” *Proc. of the IEEE*, vol. 93, no. 1, pp. 57–70, 2005.

[90] O. Campana, R. Contiero, and G. A. Mian, “An H.264/AVC video coder based on a multiple description scalar quantizer,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 18, no. 2, pp. 268–272, 2008.

[91] O. Crave, B. Pesquet-Popescu, and C. Guillemot, “Robust video coding based on multiple description scalar quantization with side information,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 20, no. 6, pp. 769–779, 2010.

[92] B. A. Heng, J. G. Apostolopoulos, and J. S. Lim, “End-to-end rate-distortion optimized MD mode selection for multiple description video coding,” *EURASIP Jnl. on Applied Signal Processing*, vol. 2006, p. 12, 2006.

[93] A. R. Reibman, H. Jafarkhani, Y. Wang, and M. Orchard, “Multiple description video using rate-distortion splitting,” in *Proc. IEEE Int’l Conf. on Image Processing*, vol. 1, 2001, pp. 978–981.

[94] C. Kim and S. Lee, "Multiple description coding of motion fields for robust video transmission," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, pp. 999–1010, 2001.

[95] A. R. Reibman, H. Jafarkhani, Y. Wnag, M. T. Orchard, and R. Puri, "Multiple description video coding using motion-compensated temporal prediction," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 3, pp. 193–204, 2002.

[96] D. Comas, R. Singh, and A. Ortega, "Rate-distortion optimization in a robust video transmission based on unbalanced multiple description coding," in *Proc. IEEE Workshop on Multimedia Signal Processing (MMSP)*, 2001, pp. 581–586.

[97] I. K. Kim and N. I. Cho, "Error resilient video coding using optimal multiple description of DCT coefficients," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 3, 2003, pp. 663–666.

[98] J. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Proc. SPIE Visual Communication and Image Processing*, 2001, pp. 392–409.

[99] S. Wenger, "Video redundancy coding in H.263+," in *Conf. on Audio-Visual Services Over Packet Networks*, 1997.

[100] V. Vaishampayan and S. John, "Balanced interframe multiple description video compression," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 3, 1999, pp. 812–816.

[101] S. Wenger, G. D. Knorr, J. Ott, and F. Kossentini, "Error resilience support in H.263+," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8, pp. 867–877, 1998.

[102] J. Apostolopoulos, "Error-resilient video compression through the use of multiple states," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 3, 2000, pp. 352–355.

[103] M. Flierl and B. Girod, "Video coding with motion-compensated lifted wavelet transforms," *Signal Processing: Image Communications*, vol. 19, no. 7, pp. 561–575, 2004.

[104] E. Akyol, A. M. Teklap, and M. R. Civanlar, "Scalable multiple description video coding with flexible number of descriptions," in *Proc. IEEE Int'l Conf. on Image Processing*, vol. 3, 2005, pp. 712–715.

[105] I. Radulovic and P. Frossard, "Balanced multiple descripiton scalar quantization," in *IEEE Int'l. Symposium on Information Theory*, 2008, pp. 1997–2001.

[106] C. Tian, J. Chen, and S. N. Diggavi, "Multiuser successive refinement and multiple description coding," *IEEE Trans. on Information Theory*, vol. 54, no. 2, pp. 921–931, 2008.

[107] T. A. Beery and R. Zamir, "Multi level multiple descriptions," in *Proc. Data Compression Conference (DCC)*, 2009, pp. 63–72.

[108] "Proposals for video coding complexity assessment," in *Joint Collaborative Team on Video Coding (JCT-VC)of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11*, 2010, pp. 1–10.

[109] Y. Shen, Z. Liu, S. S. Panwar, K. W. Ross, and Y. Wang, "Streaming layered encoded video using peers," in *Proc. IEEE Int'l. Conf. on Multimedia and Expo*, 2005, p. 4 pp.

[110] R. Rejaie and A. Ortega, "PALS: Peer-to-peer adaptive layered streaming," in *Proc. of Int'l Workshop on Network and Operating Systems Support of Digital Audio and Video (NOSSDAV)*, 2003, pp. 153–161.

[111] F. de Asis Lopez-Fuentes, "Adaptive mechanism for P2P video streaming using SVC and MDC," in *IEEE Int'l. Conf. on Complex, Intelligent and Software Intensive Systems*, 2010, pp. 457–462.

[112] P. Baccichet, T. Schierl, T. Wiegand, and B. Girod, "Low-delay peer-to-peer streaming using scalable video coding," in *Proc. Int'l. Workshop on Packet Video*, 2007, pp. 173–181.

[113] M. Mustaq and T. Ahmed, "Adaptive packet video streaming over P2P networks using active measurements," in *IEEE Symposium on Computers and Communications (ISCC)*, 2006, pp. 423–428.

[114] A. T. Nguyen, B. Li, and F. Eliassen, "Chameleon: Adaptive peer-to-peer streaming with network coding," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.

[115] E. Akyol, A. M. Teklap, and M. R. Civanlar, "A flexible multiple description coding framework for adaptive peer-to-peer video streaming," *IEEE Jnl. of Selected Topics in Signal Processing*, vol. 1, no. 2, pp. 231–245, 2007.

[116] C.-S. Lin and I.-T. Lee, "Applying multiplpe description coding to enhance the streaming scalability on CDN-P2P network," *Int'l. Jnl. of Communication Systems*, no. 23, pp. 553–568, 2010.

[117] M. H. Firooz, K. Ronasi, M. R. Pakravan, and A. N. Avanaki, "Wavelet-based unbalanced un-equivalent multiple description coding for P2P networks," in *IEEE Int'l. Conf. on Systems, Man and Cybernetics*, 2007, pp. 242–247.

[118] J. A. Pouwelse, J. R. Taal, R. L. Lagendijk, and D. H. J. Epema, "Real-time video delivery using peer-to-peer bartering netorks and multiple description coding," in *IEEE Int'l. Conf. on Systems, Man and Cybernetics*, 2004, pp. 4599–4605.

[119] J. R. Taal, J. A. Pouwelse, and R. L. Lagendijk, "Scalable multiple description coding for video distribution in P2P networks," in *Picture Coding Symposium*, 2004.

[120] X. Xu, Y. Wang, S. S. Panwar, and K. W. Ross, "A peer-to-peer video-on-demand system using multiple description coding and server diversity," in *Proc. IEEE Int'l Conf. on Image Processing*, 2004, pp. 1759–1762.

[121] L. Meng-ting, L. Chang-kuan, Y. Jason, and C. H. Homer, "Multiple description coding with spatial-hybrid interpolation for video streaming in peer-to-peer networks," *Jnl. of Zhejiang University Science A*, vol. 7, no. 5, pp. 894–899, 2006.

[122] T. B. Abnoz and A. M. Teklap, "Optimization of encoding configuration in scalable multiple description coding for rate-adaptive P2P video multicasting," in *Proc. IEEE Int'l Conf. on Image Processing*, 2009, pp. 3741–3744.

[123] E. Akyol, A. M. Teklap, and M. R. Civanlar, "Adaptive peer-to-peer video streaming with optimized flexible multiple description coding," in *Proc. IEEE Int'l Conf. on Image Processing*, 2006, pp. 725–728.

[124] M. Q. A. M. Arrifano, M. Pereira, M. M. Freire, and M. Antonini, "Robust peer-to-peer video streaming based on multiple description coding," in

IEEE Int'l Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), 2009, pp. 1–6.

[125] S. Zhu, S. A. Yeung, and B. Zeng, "A comparative study of multiple description video coding in P2P: Normal MDSQ versus flexible dead-zones," in *Proc. IEEE Int'l. Conf. on Multimedia and Expo*, 2009, pp. 1118–1121.

[126] J. R. Taal and R. L. Lagendijk, "Hybrid temporal-SNR multiple description coding for peer-to-peer television," in *Picture Coding Symposium*, 2006.

[127] C. Lamy-Bergot, B. Candillon, B. Pesquet-Popescu, and B. Gadat, "A simple multiple description coding scheme for improved peer-to-peer video distribution over mobile links," in *Picture Coding Symposium*, 2009, pp. 1–4.