Optimal Selection of Encoding Configuration for Scalable and Multiple DescriptionVideo Coding

by

Tenzile Berkin Abanoz

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of
Master of Science
in
Electrical and Computer Engineering

Koc University

August 2008

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Tenzile Berkin Abanoz

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Prof. A. Murat Tekalp (Advisor)

_____

Assist. Prof. Oğuz Sunay

_____

Assoc. Prof. Metin Türkay

Date: _____

to my family

ABSTRACT

It is well-known that the wider the range of extraction points a scalable bitstream supports, the lower the compression efficiency at these extraction points. Moreover, this compression efficiency generally varies according to what combination of scalability types are used to support this range of extraction points as specified by the encoding configuration. Hence, we propose some objective criteria as a measure of coverage, compression efficiency and rate-distortion performance of a configuration, and then present a multiple-objective optimization formulation to select the best encoding configuration for scalable video coding, given a range of bitstreams that must be supported.

Additionally, we present methods for scalable multiple description coding (SMDC) of monocular and multi-view video, where each description is scalable so that they can separately be streamed over DCCP with efficient rate adaptation. These descriptions are derived from an SVC-compliant bitstream, where there are variables related to the SVC encoding configuration and MD generation method used. We present a multiple objective optimization (MOO) framework to determine these variables so that the resulting SMDC strikes the best balance between maximizing the range of extraction points (coverage) of individual scalable descriptions, maximize the average end-to-end rate-distortion performance over the range of extraction points for a set of packet loss probabilities, and minimize the redundancy among descriptions. We performed optimization over the base layer rate (quantization parameter) and a selection of MD generation methods that feature various levels of redundancy at a fixed total rate for all descriptions. The framework is generic to allow optimization over other encoding variables as well if desired. Results of Monte-Carlo simulation of SMDC streaming of both monocular and stereoscopic videos demonstrate the performance of the proposed optimization framework.

ÖZET

Ölçeklenebilir bit katarının desteklediği özütleme noktalarının erimi ne kadar genişse, bu özütleme noktalarındaki sıkıştırma verimliliği de o kadar düşüktür. Hatta genelde sıkıştırma verimliliği kullanılan ölçeklenebilirlik türlerine göre değişir. Belirli bir bit katarı erimi içinde ölçeklenebilir video kodlama için en iyi kodlama düzenleşimini belirlemek amacıyla çoklu hedef optimizasyonu oluşturduk. Bu sebeple, kapsam, sıkıştırma verimliliği ve hız-bozunum verimi gibi bazı hedef fonksiyonları belirledik.

Ayrıca, tekgözlü ve çok bakışlı videolar için ölçeklenebilir bir çoklu betimle video kodlama öneriyoruz. Hız uyarlaması ile DCCP üzerinde aktarılabilecek şekilde, kendi içinde ölçeklenebilir özelliğe sahip her betimleme ölçeklenebilir video kodlamaya uyumlu bir bit katarından elde edilmektedir. Betimlemeler oluşturulurkenki değişkenler ölçeklenebilir video kodlama düzenleşimini ve çoklu betimleme oluşturulma yöntemini içerir. Bu değişkenleri belirlemek amacıyla bir çoklu kriter eniyilemesi öneriyoruz. Belirlenen değişkenler elde edilen ölçeklenebilir çoklu betimleme video kodlama yönteminin belirli kriterler arasında en iyi dengeyi oluşturmasını sağlıyor, bu kriterler ölçeklenebilir betimlemelerin özütleme noktalarının eriminin en büyütülmesi, bir grup paket kayıp olasılığında özütleme noktaları erimi boyunca hız-bozunum veriminin en büyütülmesi ve artıklığın en küçültülmesi. En iyilemeyi taban bit hızı (nicemleme parametresi) ve betimleme oluşturma yöntemleri üzerinden yaptık. Eniyileme modeli istenildiğinde farklı kodlama değişkenleri içerebilmeye uyumlu. Tekgözlü ve streo videolar için bulunan Monte-Carlo benzetiminin sonuçları önerilen eniyileme yönteminin başarımını göstermektedir.

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor Prof. A. Murat Tekalp, who have been a great source of inspiration and provided the right balance of suggestions, criticism, and freedom.

I am grateful to members of my thesis committee for critical reading of this thesis and for their valuable comments.

Next, I would like to thank my family who always supported me for their guidance and for their belief in me.

Also, I would like to thank my officemates, Burak Görkemli and C. Göktuğ Gürler, to my friends , Murat Çilingiroğlu A. Engin Ural, E. Pınar Karabulut, Ferda Ofli for always being there to help, and to Prof. İrşadi Aksun, Prof. Engin Erzin, E. Pınar Karabulut, Aytaç Alparslan, A. Engin Çetin, for sharing their computers to run my simulations. Finally, I would like to thank TÜBİTAK for providing the fellowship for my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

3DTV	Three Dimesional Television
HDTV	High Definition Television
SVC	Scalable Video Coding
SMDC	Scalable Multiple Description Coding
MOO	Multiple-Objective Optimization
FGS	Fine Granular Scalability
CGS	Coarse Grain Scalability
QP	Quantization Parameter
MD	Multiple Description
JVT	Joint Video Team
JSVM	Joint Scalable Video Model
MCTF	Motion-Compensated Temporal Filtering
DCCP	Datagram Congestion Control Protocol
TFRC	TCP-Friendly Rate Control
MGS	Medium Grain Scalability
MDC	Multiple Description Coding
DCT	Discrete Cosine Transform
MDSQ	Multiple Description Scalar Quantization
MDLVQ	Multiple Description Lattice Vector Quantizer
PCT	Pairwise Correlating Transform
SS-MDC	Scaling Stereo-MDC
MS-MDC	Multi-State Stereo-MDC
FEC	Forward Error Correction
NAL	Network Abstraction Layer

GOP	Group of Pictures
RRD	Redundancy Rate Distortion
SD	Single Description

Chapter 1

INTRODUCTION

1.1 Motivation

With the development of recent technologies such as 3DTV, HDTV, etc. and increase of media related transmission in the internet, there is a growing need of new frameworks for coding and transmission of media. For example, a new research project in Europe aims to pipe TV programs over the Internet, as part of P2P-Next project. Scalable Video Coding (SVC) and Scalable Multiple Description Coding (SMDC) are coding strategies which both enable rate adaptation by offering multiple extraction points. In addition to rate adaptation, SMDC also enables path diversity.

Both of these video coding strategies require an input configuration file containing several parameters that offer tradeoff between video quality and rate for varying network conditions. This thesis proposes Multiple-Objective Optimization (MOO) of the encoding configuration for both SVC and SMDC with more than one, possibly conflicting optimization criteria.

1.2 Problem Formulation

SVC includes multiple scalability types, such as temporal scalability, spatial scalability, and two quality scalability choices: Coarse Grain Scalability (CGS) or Medium Grain Scalability (MGS). We can combine these scalability types or use them separately. In

addition, MPEG-4 Part 2 video coding standard supports another quality scalability option, called Fine Grain Scalability (FGS). The number of layers can change for each of these scalability types. Additionally, quantization parameter (QP) value for each of the layers generated by these scalability types is an important parameter we need to consider in our optimization.

Chapter 2 addresses the problem of selection of the best encoding configuration for MPEG4 part 2 scalable video coding, namely scalability type and number of layers for each type, and quantization parameter for base FGS layer at each spatial resolution in order to meet some conflicting criteria including maximization of the coverage of extraction points within a predetermined bitrate range, maximization of the rate-distortion performance at each extraction point within the given range, maximization of the average incremental compression efficiency of each layer over all extraction points within the range, and maximization of the maximum picture size represented by the encoder configuration within the given bitrate range.

In Chapter 3, we propose a scalable multiple description coding method based on SVC using MGS layers with different encoding configuration parameters, QP for base layer, and multiple MD generation methods for both monocular and multi-view video. The MD generation methods rely on the concept of including the base layer as redundancy in all descriptions and splitting MGS NAL units of different frames to different descriptions. Then, we propose a MOO formulation for optimization of these parameters to find the configuration that strikes the best balance between coding efficiency, rate-distortion performance, and the redundant bits in both descriptions considering the network loss probabilities.

Chapter 4 concludes the results of MOO formulation for SVC and SMDC, and offers some future work.

1.3 Contributions

So far the configuration file for scalable video encoding was prepared in an ad-hoc manner, in chapter 2 we offer a novel optimization framework considering different objective criteria for selecting the encoding configuration. This work is presented in the IEEE International Conference on Image Processing 2007.

Chapter 3 offers an optimization framework similar to Chapter 2, but for scalable multiple description coding. Additionally, it presents methods for generating individually scalable descriptions, which enables TFRC rate adaptation through DCCP and which enables path diversity. There is no previous work based on H.264/AVC which generates individually scalable bitstreams. Also, it analyzes the effect of different objective criteria and scaling methods on MOO results. The work on this chapter is in the process of submission as a journal paper.

Chapter 2

OPTIMAL SELECTION OF ENCODING CONFIGURATION FOR SCALABLE VIDEO CODING

2.1 Introduction

Standardization of scalable video coding (SVC), which is often desirable for efficient rate adaptation in video transport over the Internet, is a current work item under the Joint Video Team (JVT) [1]. The reference encoder-decoder, called the Joint Scalable Video Model (JSVM) [2], is based on a scalable extension of the well-established JVT standard H.264/AVC. It provides temporal scalability using of motion-compensated temporal filtering (MCTF) implemented by a lifting framework. For spatial scalability, a combination of motion-compensated prediction and over-sampled pyramid decomposition is employed [3]. SNR scalability is achieved by residual quantization with some modification to the H.264/AVC syntax.

While temporal and spatial scalability modes of SVC allow bitstream extraction at specific rate-distortion points, the FGS mode allows extraction more-or-less over a continuous range of rate-distortion points. The concept of quality layers has been introduced in [4] to allow for rate-distortion optimized bitstream extraction over a range of rates. However, the number and range of extraction points is determined by the encoder configuration, which is often determined prior to encoding in an ad-hoc manner. In order to have flexibility for adaptation of the video rate to a wide range of network conditions, we

would like to have as many extraction points within a predetermined operating range of bitrates, which requires definition of several scalability layers. However, the number and type of scalability layers may have significant impact on the compression efficiency, and cannot be easily optimized over a wide range of bitrates.

This chapter addresses the problem of selection of the best encoding configuration, namely scalability type and number of layers for each type, and quantization parameter for base FGS layer at each spatial resolution in order to meet some conflicting criteria including maximization of the coverage of extraction points within a predetermined bitrate range, maximization of the rate-distortion performance at each extraction point within the given range, maximization of the average incremental compression efficiency of each layer over all extraction points within the range, and maximization of the maximum picture size represented by the encoder configuration within the given bitrate range. Mathematical definition of these criteria is given in Section 2.2. We pose a multiple objective optimization (MOO) formulation and provide a solution to this problem in Section 2.3. Experimental results are presented in Section 2.4, and conclusions are reached in Section 2.5.

2.2 Definition of Objective Criteria

In this section, we propose three criteria to quantify an encoding configuration.

2.2.1 Coverage of a Configuration

Given a target bitrate range, we define total coverage C of a scalable bitstream as how much of this range is actually covered by extraction points of this bitstream. Clearly, total coverage is the sum of coverage of the individual layers, $c(i)$, where spatial layers provide single extraction points, and FGS layers provide a range of points. Hence,

$$C = \sum_{i=1}^{L} c(i) \qquad (2.1)$$

$$c(i) = \begin{cases} 1 & \text{if } i \text{ is a spatial scalability layer} \\ (PSNR_{max} - PSNR_{min})/0.2 & \text{if } i \text{ is an FGS layer} \end{cases}$$

$PSNR_{max}$ and $PSNR_{min}$ denote the maximum and minimum PSNR values for an FGS layer calculated after interpolating to 4cif resolution, and L is the number of layers within the bitrate range. We used the PSNR value 0.2 because it is the PSNR difference which can be perceived by human eye. Additionally, the importance of coverage increase from 25 to 30 is not as important as coverage increase from 10 to 30. For this reason, we propose as objective criteria base three logarithm of the coverage function. The other base logarithms are also tried and they gave the same results as base three logarithm.

2.2.2 Efficiency of a Configuration

"Base layer usage" has been proposed as a measure of efficiency of scalable video coding in case of a base layer and one enhancement layer [5]. It measures the efficiency of scalable coding compared to simulcasting at corresponding two rates, given by

$$B = \frac{(R_S + R_B) - R_E}{R_B} = 1 - \frac{R_E - R_S}{R_B} \qquad (2.2)$$

where R_B, R_E and R_S stand for the base layer rate, total rate of scalable stream (base + enhancement), and the simulcast (non-scalable coding) rate at the same PSNR as that of total scalable rate, respectively. B takes values in the range 0-1. When $B \approx 0$, the efficiency

of scalable coding is close to simulcasting, and when $B \approx 1$, the efficiency of scalable coding is close to non-scalable coding at the same quality, which is desirable.

We hereby extend this definition to the case of more than one enhancement layers, called incremental efficiency, *b(i)*, of layer *i*. The *b(i)* is defined as in (2), except that the base layer rate is taken as the total rate of all layers up to layer *i*. This is because all those layers, including layer *i*-1, are required for decoding enhancement layer *i*.

$$b(i) = \frac{R_s(i) + R_E(i-1) - R_E(i)}{R_E(i-1)} = 1 - \frac{R_E(i) - R_S(i)}{R_E(i-1)} \qquad (2.3)$$

where $R_S(i)$ is the single layer rate giving the PSNR value of layer I, $R_E(i-1)$ is the total rate of the scalable stream including *i-1* layers.

Then, we define the overall efficiency, *E*, of a scalable bitstream as the average of incremental efficiency of all layers that fall within bitrate range of interest, given by

$$E = \frac{\sum_{i=1}^{L} b(i)}{L} \qquad (2.4)$$

where *L* is the number of layers within the bitrate range.

2.2.3 Rate-Distortion Performance of a Configuration

It is desirable that we have good rate-distortion (RD) performance at all possible extraction points of a scalable bitstream. Hence, we define the overall RD performance of a scalable bitstream as the average of RD performances over all possible extraction points, given by

$$RD = \frac{\sum_{i=1}^{C} [\lambda_i R(i) + D(i)]}{C} \qquad (2.5)$$

$\lambda_i = 0.85 * 2\text{^} ((QP_i - 12)/3)$ [6], where QP_i is the quantization parameter for layer i, $R(i)$ is the bitrate for layer i and $D(i)$ is the SSD distortion for layer i, given by

$$D(i) = \frac{255^2 * width * height}{10^{PSNR(i)/10}} \qquad (2.6)$$

and PSNR(i) is the PSNR of layer i after it is interpolated to 4cif resolution.

2.3 Multiple-Objective Optimization Formulation

Multiple objective optimization (MOO) was introduced by Pareto for solution of an optimization problem with P possibly conflicting objective functions $f_1, f_2, \ldots, f_P$. A solution s^* is called globally Pareto-optimal if any one of the objective function values cannot be improved without degrading other objective values. Then, a Pareto-optimal solution s^* exists if there exists no other feasible solution s that satisfies

$$f_p(s) \le f_p(s^*), \quad \forall p \in \{1,\ldots,P\} \qquad (2.7)$$

with at least one strict inequality. Since different objective functions represent different aspects of the problem, it is difficult to discriminate between these Pareto-optimal points and determine which one is better than the other. The MOO defines a so called *best compromise solution* as the feasible solution that is closest to the utopia point, which is an infeasible solution obtained by minimizing each objective individually [7].

In our problem, there are P=4 optimization criteria. In addition to the base three logarithm of coverage, efficiency, and rate-distortion performance criteria, which are defined in Section 2, we also employ maximum picture size as an optimization criterion, since video with the largest size should be preferred if all other parameters were equal. These criteria are optimized with respect to the encoder configuration parameters, which are type (spatial and/or FGS) and number L of scalability layers, and the quantization parameter QP for the base FGS layer at each spatial resolution.

We perform multiple-objective optimization subject to a maximum rate constraint. That is, if there is a scalable bitstream whose total bitrate is greater than the maximum target bitrate, we discard those layers with the minimum bitrate greater than the maximum target bitrate in defining feasible solutions.

Our objective is to find the encoding configuration j that strikes the best balance between

- Maximize coverage

$$\max_j \left(\log_3 (C_j)\right) \tag{2.8}$$

where C_j is the coverage for bitstream j, given by (2.1);

- Maximize efficiency

$$\max_j \left(E_j\right) \tag{2.9}$$

where E_j is the overall efficiency of bitstream j, given by (2.3);

- Minimize rate-distortion performance

$$\min_j \left(RD_j\right) \tag{2.10}$$

where RD_j is the rate distortion for bitstream j given by (2.4);

- Maximize the maximum picture size

$$\max_j \left(\max\, p_j\right) \tag{2.11}$$

where max p_j is the maximum picture size for bitstream j.

We assume that all objectives have equal importance; hence, their values will be scaled to range [0,1] as

$$f_{scaled} = \frac{f - f_{\min}}{f_{\max} - f_{\min}} \tag{2.12}$$

where f is the original objective value prior to scaling.

In order to find the bitstream that strikes the best compromise between our four optimization criteria, we first encode the video with N different choices of encoding configurations. Each one of these encoding configurations is a feasible solution point. The final step is to find the solution point that is closest to the (infeasible) utopia point which is the point (0,1,1,1) after normalization. The utopia point is the one where all optimization criteria are satisfied, in other words, where rate-distortion measure is minimum, and maximum picture size, coverage and coding efficiency are all maximum. To find the closest feasible point, we use Euclidian distance, and calculate the distance for each configuration j as

$$d_j = \sqrt{\left(1 - E_j\right)^2 + \left(1 - \max p_j\right)^2 + \left(1 - \log_3(C_j)\right)^2 + \left(RD_j\right)^2} \tag{2.12}$$

and we select the configuration with minimum distance $\min_i(d_i)$ as the MOO solution.

2.4 Experimental Results

We consider the following problem: Design a scalable bitstream that should simultaneously serve i) a broadband client at about 1.5-3 Mbps, ii) a DSL client at near 256 kbps. To this effect, we encoded two video files (soccer.yuv, harbour.yuv) with *N*=21 different encoding configurations, that is, with different number of spatial and FGS layers, and with different quantization parameters. The list of feasible configurations is shown in Table 2.1 for the video "soccer.yuv" and in Table 2.2 for the video "harbour.yuv". The configurations are specified as base spatial layer-followed by the number of FGS layers at that spatial resolution-followed by quantization parameter for the base FGS layer at that spatial resolution. The + sign indicates the next spatial layer specified in the same format. For example, cif-2-38 + 4cif-2-38 denotes that base spatial layer is cif and we have 2 FGS layers for cif resolution, and the quantization parameter for base FGS layer is 38, followed by 4cif resolution with the same parameters.

The values of coverage, base three logarithm of coverage, efficiency, maximum picture size rate-distortion and distance values for each configuration for both of the videos are listed in the Table 2.1 and Table 2.2. The distance of each configuration as listed in Table 2.1 and Table 2.2 to the utopia point is plotted in Figure 2.1.

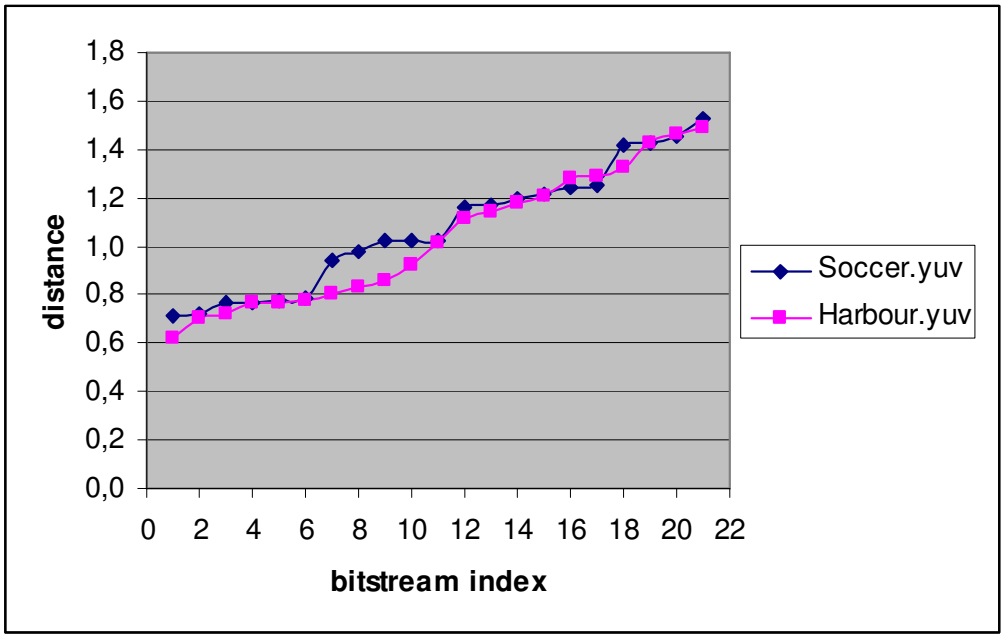


Fig. 2.1: Distance of configurations to the utopia point by rank

2.5 Discussions

Inspection of Table 2.1 and Table 2.2 shows that the first five configurations for both videos include the same four configurations, namely cif-1-38 + 4cif-1-38; cif -2-38 + 4cif-2-38; cif-1-40 + 4cif-1-40; cif -2-40 + 4cif-2-40, which indicates that these are the best encoder configurations Their actual rankings are somewhat different, because the two videos contain different amount of spatial and temporal detail.

If we analyze the properties of the first two ranked configurations for “soccer.yuv”, we see that their distances are close to each other, but the efficiency of the first configuration is better than that of the second while their coverage and RD values are close to each other.

We note that the best solution may change if the users express different preference (weights) on different criteria.

Table 2.1: Ranking of several encoding configurations according to their normalized Euclidean distance to the utopia point.

Soccer.yuv							
Configuration Parameters	**Conf. Effici.**	**Max. Pic. Size**	**Conf. Cov.**	**Log3(Cov.)**	**Conf. RD**	**Distance**	**Rank**
cif-1-38 + 4cif-1-38	0,674	405504	21	2,771	16729169	0,26	1
cif -2-38 + 4cif-2-38	0,619	405504	27	3,000	16550397	0,27	2
cif -2-40 + 4cif-2-40	0,616	405504	32	3,155	18256597	0,31	3
Qcif- 0-32 + cif-0-32 + 4cif-2-40	0,624	405504	18	2,631	13359059	0,32	4
cif-1-40 + 4cif-1-40	0,670	405504	21	2,771	20183890	0,37	5
Qcif- 0-40 + cif-0-40 + 4cif-2-40	0,570	405504	22	2,814	14772208	0,40	6
Qcif-0-32 + cif-1-40 + 4cif-2-40	0,567	405504	26	2,966	17701892	0,43	7
Qcif-1-38 + cif-1-38 + 4cif-1-38	0,555	405504	28	3,033	21972394	0,56	8
Qcif-1-34 + cif-1-34 + 4cif-1-34	0,590	405504	16	2,524	21915195	0,56	9
Qcif-1-40 + cif-1-40 + 4cif-1-40	0,549	405504	27	3,000	25357453	0,68	10
Qcif-2-38 + cif-2-38 + 4cif-2-38	0,494	405504	30	3,096	25157722	0,79	11
Qcif- 2-40 + cif-2-40 + 4cif-2-40	0,472	405504	39	3,335	24679739	0,82	12
Qcif-1-32 + cif-1-32 + 4cif-1-32	0,654	101376	12	2,262	23597170	1,17	13
Qcif-2-38 + cif-2-38	0,533	101376	25	2,930	27835771	1,28	14
Qcif-0-32 + cif-0-32 + 4cif-0-32	0,371	405504	3	1,000	20820273	1,34	15
Qcif-1-38 + cif-1-38	0,618	101376	15	2,465	31578156	1,34	16
Qcif-2-40 + cif-2-40	0,516	101376	26	2,966	30784696	1,37	17
Qcif-0-34 + cif-0-34 + 4cif-0-34	0,360	405504	3	1,000	23309491	1,39	18
Qcif-1-40 + cif-1-40	0,598	101376	14	2,402	35915953	1,48	19
Qcif-0-32 + cif-0-32	0,682	101376	2	0,631	27654914	1,55	20
Qcif-0-34 + cif-0-34	0,677	101376	2	0,631	30437866	1,60	21

Table 2.2: Ranking of several encoding configurations according to their normalized Euclidean distance to the utopia point.

Harbour.yuv							
Configuration Parameters	**Conf. Effici.**	**Max. Pic. Size**	**Conf. Cov.**	**Log3(Cov.)**	**Conf. RD**	**Distance**	**Rank**
cif -2-38 + 4cif-2-38	0,567	405504	24	2,893	35091883	0,27	1
cif-1-40 + 4cif-1-40	0,563	405504	24	2,893	36246351	0,28	2
cif-1-38 + 4cif-1-38	0,538	405504	25	2,930	30209428	0,29	3
Qcif- 0-40 + cif-0-40 + 4cif-2-40	0,538	405504	22	2,814	29226375	0,31	4
cif -2-40 + 4cif-2-40	0,520	405504	30	3,096	36216258	0,34	5
Qcif-1-38 + cif-1-38 + 4cif-1-38	0,482	405504	28	3,033	45981028	0,50	6
Qcif-1-40 + cif-1-40 + 4cif-1-40	0,505	405504	29	3,065	56904820	0,58	7
Qcif-0-32 + cif-1-40 + 4cif-2-40	0,384	405504	28	3,033	36018061	0,65	8
Qcif- 0-32 + cif-0-32 + 4cif-2-40	0,380	405504	18	2,631	25779420	0,68	9
Qcif-2-38 + cif-2-38 + 4cif-2-38	0,443	405504	25	2,930	62524322	0,74	10
Qcif-1-34 + cif-1-34 + 4cif-1-34	0,459	405504	15	2,465	61130323	0,75	11
Qcif- 2-40 + cif-2-40 + 4cif-2-40	0,421	405504	35	3,236	62316472	0,77	12
Qcif-1-32 + cif-1-32 + 4cif-1-32	0,515	101376	11	2,183	60294933	1,24	13
Qcif-2-38 + cif-2-38	0,496	101376	22	2,814	68342409	1,25	14
Qcif-0-34 + cif-0-34 + 4cif-0-34	0,343	405504	3	1,000	66496759	1,28	15
Qcif-1-38 + cif-1-38	0,612	101376	13	2,335	78193351	1,32	16
Qcif-2-40 + cif-2-40	0,498	101376	25	2,930	78702638	1,33	17
Qcif-1-40 + cif-1-40	0,645	101376	15	2,465	88819162	1,41	18
Qcif-0-32 + cif-0-32 + 4cif-0-32	0,227	405504	3	1,000	62118924	1,43	19
Qcif-0-34 + cif-0-34	0,597	101376	2	0,631	92271302	1,74	20
Qcif-0-32 + cif-0-32	0,445	101376	2	0,631	87305844	1,76	21

Chapter 3

OPTIMIZATION OF SCALABLE MULTIPLE-DESCRIPTION CODING FOR MONOCULAR AND MULTI-VIEW VIDEO

3.1 Introduction

Vast majority of the Internet traffic is already media-related and this trend continues to strengthen in the future. Hence, there is need for new streaming strategies, such as peer-to-peer and content-aware networks with congestion control, and new streaming protocols and video coding methods that are suitable for such schemes.

A recent protocol for media streaming with congestion control is the DCCP featuring TCP-friendly rate control (TFRC) [8, 9]. Scalable video coding (SVC) provides a natural means for efficient rate-adaptation to match the TFRC rate for video transport over DCCP. Standardization of SVC is in its final stages under the Joint Video Team (JVT) [10]. The reference encoder-decoder, called the Joint Scalable Video Model (JSVM) [11], is based on a scalable extension of the well-established JVT standard H.264/AVC. Medium grain scalability (MGS) is one of the scalability types in the current JSVM, which is based on splitting the sequences of transform coefficients into a certain number of fragments, for each 4x4 or 8x8 block [12].

Multiple description coding (MDC) is a source coding technique, which encodes video in multiple independently decodable streams, called descriptions, at the expense of introducing some redundancy. It enables decoding of video in a lower quality if some of

the descriptions are received, and in full quality if all descriptions are received. There are many works on MDC which are reviewed in [13]. MDC offers a robust solution for video delivery over unreliable networks and peer-to peer streaming if it is combined with path/server diversity [14].

An early approach to create multiple descriptions is multiple description scalar quantization (MDSQ), where redundancy is introduced in the quantization step. In this approach, two coarse quantizers with overlapping cells are used [15, 16]. The rate analysis of MDSQ is presented in [17]. Trellis coded quantization [18] and multiple description lattice vector quantizer are proposed (MDLVQ) [19] as improvements over MDSQ. Further MD codes based on MDSQ are proposed in [20, 21, 22]. Additional MDLVQ coders are proposed in [23, 24, 25].

Another MD approach is based on using pairwise correlating transform (PCT) [26], where the input signal is first decorrelated with a proper transform, for example DCT, and are coupled in pairs. Then, by the PCT two decorrelated coefficients become two correlated coefficients. Whereas one coefficient is sent in Description 1, the other is sent in Description 2. MDC schemes using this method are proposed in [27, 28, 29, 30].

Additionally, MDC with multiple prediction loops in encoder to overcome the mismatch between the encoder and the decoder is introduced in [31, 32, 33, 15]. Motion compensated multiple descriptions coding which is motion estimation across descriptions is proposed in [34]. A balanced MDC has been proposed in [35] by splitting the DCT transform coefficients in non-scalable bitstreams. In [36], an unbalanced multiple description video coding, where descriptions with unequal rates are created, is introduced.

Multiple description scalable coding using wavelet based motion compensated temporal filtering is introduced in [37], combining the advantages of MDC and SVC. Another multiple description scalable coding based on wavelet transform [38] introduces a flexible multiple description coding framework.

An overview of MDC and its relevance to 3DTV is presented in [39]. [40] presents and compares two non scalable MDC framework for stereoscopic video, Scaling Stereo-MDC (SS-MDC) and Multi-State Stereo-MDC (MS-MDC), which are based on H.264.

While SVC offers packet loss resilience and means for efficient rate adaptation, MDC offers packet loss resilience and path diversity. Therefore, a combination of SVC and MDC should offer the highest flexibility in streaming video over the Internet. This combination may be achieved by: i) Hybrid scalable multiple description coding [41], where the base layer is streamed separately over a proper forward error protection (FEC) channel, and multiple enhancement layer descriptions are generated. ii) Layered multiple descriptions with separate base and enhancement layer descriptions that contain various levels of FEC [42, 43, 44, 45], where base layer descriptions can be streamed to low bandwidth users and both base and enhancement descriptions can be streamed to high bandwidth users. iii) Multiple descriptions, where each description is scalable such that each description can be streamed over DCCP with effective rate adaptation using potentially different links.

This chapter focuses on the last approach, and proposes methods for generation of multiple scalable descriptions from an SVC bitstream with MGS layers. Since there are a variety of ways to generate multiple descriptions from a pre-encoded SVC bitstream and multiple parameters in the SVC encoding configuration, we propose a multiple-objective optimization (MOO) formulation, which optimizes the rate-distortion objective of all descriptions, as well as the range of extraction points (coverage of the SMDC) and overall redundancy in the MDC in order to select the best SMDC configuration. Section 3.2 introduces the proposed framework for scalable MD optimization, including various SMDC generation methods, the optimization criteria and variables, as well as the MOO formulation for selection of the best SMDC encoding configuration for the case of monocular video. Section 3.3 extends this formulation to SMDC of stereo video.

Experimental results for both mono and stereo video are presented in Section 3.4. Section 3.5 presents some conclusions.

3.2 Framework for Optimization of Scalable Multiple-Description Coding

Scalable multiple description coding aims to simultaneously serve clients with different bandwidths and quality requirements. Since there are many ways to generate scalable multiple descriptions from a pre-encoded SVC bitstream, and multiple parameters in the SVC encoding configuration, selecting the "best" SMDC encoding strategy to maximize the coding efficiency, range of extraction points, as well as to minimize the redundancy becomes an important design problem in preparation of content for adaptive content-aware streaming. This section proposes an off-line multiple-objective optimization (MOO) formulation to determine the best MD generation and SVC encoding configurations for a given video. Some candidate quality scalable MD generation methods are presented in Section 3.2.1. Section 3.2.2 presents three objective criteria; namely, mean rate-distortion performance, range (coverage) of extraction points, and overall redundancy. The MOO formulation is given in Section 3.2.3. The *variables* of the optimization include the SVC encoding configuration parameters and candidate MDC generation methods. Clearly, the proposed optimization framework is not limited only to those MD generation schemes described in Section 3.2.1, but can be used with any other MD generation method.

3.2.1 Candidate Scalable MD Generation Methods

In the following, we describe five candidate MD generation methods, each creating two descriptions based on quality scalability, as possible candidates for optimization. They provide a tradeoff between the overall redundancy and rate-distortion performance of each

description. The common concept used by these methods is to include base layer in both descriptions and split the MGS NAL units between the descriptions based on a periodic pattern of group of pictures (GOP). Since MGS scalability uses the concept of hierarchical coding, odd frames uses even frames as reference. If we assign MGS NAL units of odd frames to one description and MGS NAL units of even frames to the other, the rate and PSNR values may be unbalanced. Therefore, we use a period of two group of pictures and each description will include MGS NAL units of even frames for a GOP, and MGS NAL units of odd frames for the next GOP. The composition of description 1 and description 2 for five different MD generation methods are depicted in Figures 3.1 and 3.2, respectively. In all methods, "B" denotes that the frame with number shown below contains just the base layer, "B+E" indicates that the frame contains both the base layer and all MGS layers. The meaning of "B+Ex" varies in the different methods, indicating varying amount of redundancy, which is explained below:

Method I: "B+Ex" is identical to "B+E" indicating that the frame contains both base layer and all MGS layers. That is, both descriptions include base layer NAL units for all frames and all MGS NAL units for all frames at the beginning of each GOP.

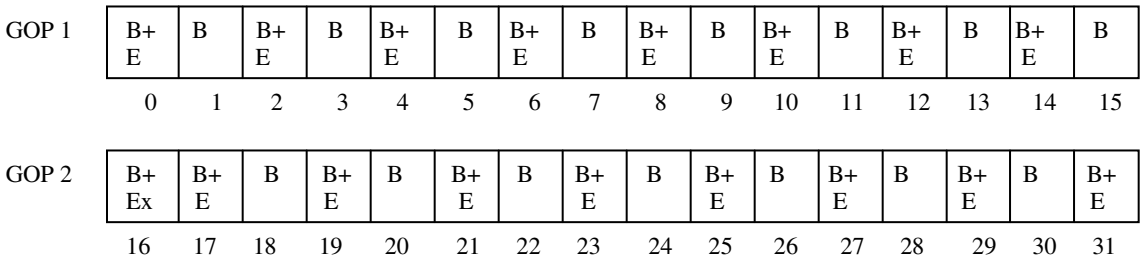
Method II: "B+Ex" indicates base layer plus MGS NAL unit containing just the DC coefficient. That is, Method II has less redundancy compared to Method I.

Method III: "B+Ex" stands for base layer plus MGS NAL units for the first four transform coefficients. That is, the redundancy of Method III is between those of Method I and Method II.

Method IV: This method is similar to Method I, except that the odd frames are completely omitted from both descriptions if they contain only the base layer "B", since they are not used as reference frames. Referring to Figures 3.1 and 3.2, for the first description frames 1,3,5,….,13,15 are omitted and for the second description frames 17, 19,

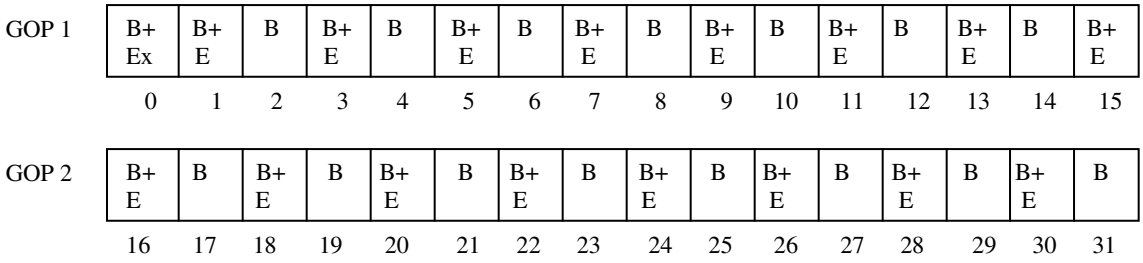
21, …, 29, 31 are omitted. Note that, this method can be considered as a modified version of the simple MD method, where even and odd frames are split to different descriptions.

Method V: “B+Ex” is identical to “B” the base layer. That is, the redundancy of this method is equal to only the base layer bits.



Description 1

Figure 3.1: Scalable MDC Description 1



Description 2

Figure 3.2: Scalable MDC Description 2

The quantization parameter of the base layer determines the rate of the base layer. When we increase the rate of the base layer but fix the quantization parameter value for the enhancement layer, the redundancy also increases, but we guarantee a better visual quality and since the difference between the base and the enhancement layer decreases, we expect a better rate distortion performance for the encoding configuration.

3.2.2 Definition of Objective Criteria

Mean Rate-Distortion Objective: Since one of the benefits of scalable and multiple description coding is providing error-resilience, we will compute the rate-distortion objective of a particular configuration considering the distortion of the received description streams over a packet-loss network. Given a description j at the extraction point i of a particular encoding configuration, we define the sum square difference (SSD) distortion $D_j(i)$ by

$$D_j(i) = \frac{255^2 \text{ W H}}{10^{PSNR_j(i)/10}} \tag{3.1}$$

where W and H denote the width and height of a frame, and $PSNR_j(i)$ is the peak signal to noise ratio of the received description over a packet loss network with the packet loss probability p. For the case of two descriptions, the index j can be 1, 2, or 3, where index 3 indicates the central distortion, that is, the combination of both descriptions. In our experiments, the packet losses will be simulated by means of the Monte Carlo method. We define the mean distortion value over N_{MC} Monte Carlo simulations as follows

$$\bar{D}_j(i) = \frac{1}{N_{MC}} \sum_{n=1}^{N_{MC}} D_j^n(i) \tag{3.2}$$

where $D_j^n(i)$ denotes the SSD distortion of the n'th Monte-Carlo simulation.

Furthermore, it is desirable that we have good rate-distortion (RD) performance over all possible extraction points of a scalable bitstream. Hence, we define the mean rate-distortion objective of a scalable description as the average of RD values over a set of sample extraction points as follows

$$RD_j = \frac{\sum_{i=1}^{N_{EP}} \left[\bar{D}_j(i) + \lambda_i R_j(i) \right]}{N_{EP}} \tag{3.3}$$

where N_{EP} is the number of extraction points, $R_j(i)$ is the bitrate for extraction point i, $\lambda_i = 0.85 * 2^{((QP_i - 12)/3)}$ and QP_i is the quantization parameter for the layer that extraction point i belongs to.

We need to compute the mean rate-distortion objective of each description j for N different encoding configurations, denoted by $RD_j(k)$, k=1,…, N. We finally normalize them to the range [0,1] by

$$NRD_j(k) = \frac{\min\{RD_j(1),\ldots,RD_j(N)\}}{RD_j(k)} \tag{3.4}$$

Range of Extraction Points: As the bitrate range covered by the extraction points increases, the rate of the bitstream can be better adapted to varying network rate. We define coverage of a description j of configuration k as the bitrate range of allowed extraction points

$$C_j(k) = MaxBR_j(k) - MinBR_j(k) \tag{3.5}$$

where $MaxBR_j(k)$ and $MinBRj(k)$ stands for maximum and minimum bitrates for description j of configuration k.

Redundancy: The redundancy of an MD coder at a particular rate-distortion operating point has been quantified by the Redundancy-Rate Distortion (RRD) curve [27]. If R^* is the rate of a single description (SD) coder when the distortion is D_0 and if R is the rate when the central distortion of the MD coder is D_0, then the redundancy is defined by

$$\rho = R - R^* \tag{3.6}$$

In our formulation, we define the relative redundancy of an encoding configuration k by

$$RR(k) = \frac{\rho(k)}{R^*(k)} \tag{3.7}$$

where $RR(k)$ is in the range [0,1] by definition.

3.2.3 Multiple Objective Optimization Formulation

In order to select the best SMDC configuration over a set of N possible SMDC encoding configurations, we propose a multiple-objective optimization (MOO) formulation with P=7 objectives, which are the mean rate-distortion objective when description 1 is received (RD_1 or NRD_1), when description 2 is received (RD_2 or NRD_2), and when both descriptions are received (RD_3 or NRD_3), the coverage when description 1 is received (C_1), when description 2 is received (C_2), and when both descriptions are received (C_3), and the relative redundancy (RR). The optimization is performed with respect to a set of candidate MD generation methods and the quantization parameter (QP) of the base layer when the QP for enhancement layer is fixed.

In order to find the configuration k that strikes the best compromise between our seven objectives, we first generate N different SMDC bitstreams, each with 2 descriptions, according to a set of N preselected encoding configurations. Each one of these SMDC bitstreams is a feasible solution point according to MOO terminology (see Section 2.3). Next, we define the utopia point u^* as an infeasible point in the solution space, where all objectives are optimized simultaneously. The final step is to find the feasible solution point that is closest to this infeasible utopia point according to some distance measure. In this work, we use the Euclidian distance of each configuration k, $k=1,\ldots, N$, to the utopia point. Since the units and dynamic range of each objective may be different, the commonly

employed approach in the MOO literature is to normalize each objective independently to the range [0,1] before computing the distance function.

The normalization can be performed either by simultaneously stretching the minimum value of an objective to 0 and the maximum value to 1 (see Eqn. 2.11), or by taking the ratio of each objective value to the maximum value of that objective, or by taking the inverse of the ratio of each objective value to the minimum value of that objective.

Since there are three rate-distortion and coverage objectives and one redundancy objective for each configuration, we scaled the three coverage and rate-distortion objectives by 1/3 each after the normalization, so that rate-distortion, coverage and redundancy objectives compete against each other on equal basis.

In the following, we describe some possible alternative methods for the computation of the distance function using combinations of the above normalization options:

Method 1: We first compute the NRD values given by equation 3.4, and then NRD_1, NRD_2, NRD_3, and redundancy RR values are rescaled to the range [0,1] by equation 2.11. In addition, the coverage values C_1, C_2, and C_3 are normalized to the range [0,1] also using Eqn. 2.11. Therefore, the utopia point is given by u^*=(⅓,⅓,⅓,⅓,⅓,⅓,0) and the distance $d_1(k)$ to be minimized is defined by

$$d_1(k)=\sqrt{\left(\frac{1}{3}-NRD_1(k)\right)^2+\left(\frac{1}{3}-NRD_2(k)\right)^2+\left(\frac{1}{3}-NRD_3(k)\right)^2+\left(\frac{1}{3}-C_1(k)\right)^2+\left(\frac{1}{3}-C_2(k)\right)^2+\left(\frac{1}{3}-C_3(k)\right)^2+(RR(k))^2} \quad (3.8)$$

Method 2: RD values are not normalized by Eqn 3.4, but all of RD, coverage and redundancy values are scaled to the range [0,1] by equation 2.11. Hence, the utopia point is now given by u^*=(0,0,0,⅓,⅓,⅓,0) and the distance $d_2(k)$ to be minimized is defined by

$$d_2(k)=\sqrt{(RD_1(k))^2+(RD_2(k))^2+(RD_3(k))^2+\left(\frac{1}{3}-C_1(k)\right)^2+\left(\frac{1}{3}-C_2(k)\right)^2+\left(\frac{1}{3}-C_3(k)\right)^2+(RR(k))^2} \quad (3.9)$$

Method 3: RD vales are normalized by Eqn. 3.4 and coverage values are normalized by

$$C_j(k) = \frac{C_j(k)}{\max\{C_j(1),\ldots,C_j(N)\}} \tag{3.10}$$

The redundancy values are not rescaled. Therefore, the utopia point is given by u^*=(⅓,⅓,⅓,⅓,⅓,⅓,*minRR*) and the distance $d_3(k)$ to be minimized is defined by

$$d_3(k) = \sqrt{\left(\frac{1}{3} - NRD_1(k)\right)^2 + \left(\frac{1}{3} - NRD_2(k)\right)^2 + \left(\frac{1}{3} - NRD_3(k)\right)^2 + \left(\frac{1}{3} - C_1(k)\right)^2 + \left(\frac{1}{3} - C_2(k)\right)^2 + \left(\frac{1}{3} - C_3(k)\right)^2 + (\min RR - RR(k))^2} \tag{3.11}$$

where *minRR* stands for the minimum relative redundancy value among the set of *N* candidate configurations.

Method 4: RD values and redundancy values are normalized by equation 3.4 and coverage values by equation 3.10. Therefore, the utopia point is given by u^*=(⅓,⅓,⅓,⅓,⅓,⅓,1) and the distance $d_4(k)$ to be minimized is defined by

$$d_4(k) = \sqrt{\left(\frac{1}{3} - NRD_1(k)\right)^2 + \left(\frac{1}{3} - NRD_2(k)\right)^2 + \left(\frac{1}{3} - NRD_3(k)\right)^2 + \left(\frac{1}{3} - C_1(k)\right)^2 + \left(\frac{1}{3} - C_2(k)\right)^2 + \left(\frac{1}{3} - C_3(k)\right)^2 + (1 - RR(k))^2} \tag{3.12}$$

The effect of these different normalization and rescaling options on the MOO results are analyzed in the experimental results given in Section 3.4.

3.3. Extension to Scalable MDC of Stereo Video

Multi View Scalable Multi-Description Coding (MV-SMDC) can serve needs of different clients who may need different views of a video. For example, if we have a video with nine views, a client with a lenticular display may require all nine views whereas a client with a stereo display may require only two of these views with a fixed view angle. Additionally, scalability of each view enables adaptation to dynamic bandwidth variations by offering multiple extraction points at different rates. Moreover, multiple description

property secures access to the video at a lower quality in case of high packet losses. In the following, we first introduce some candidate methods for generating SMDC bitstreams for a stereo video. We then present the MOO formulation to select the best configuration.

3.3.1 Scalable MD Generation Methods for Stereo Video

SMDC descriptions for stereo video can be generated by two general approaches. The first approach employs one of the five monocular MD generation methods described in Section 3.2.1 to generate descriptions 1 and 2 for each of view 1 and view 2 independently. Then, description 1 for view 1 and description 2 for view 2 can be coupled to form description 1 of stereo video (see Figure 3.3), and description 2 for view 1 and description 1 for view 2 can be coupled to form the second stereo description (see Figure 3.4). Since there are 5 different monocular MD generation methods in Section 3.2.1, this approach can be used to generate 5 different candidate stereo MD generation methods.

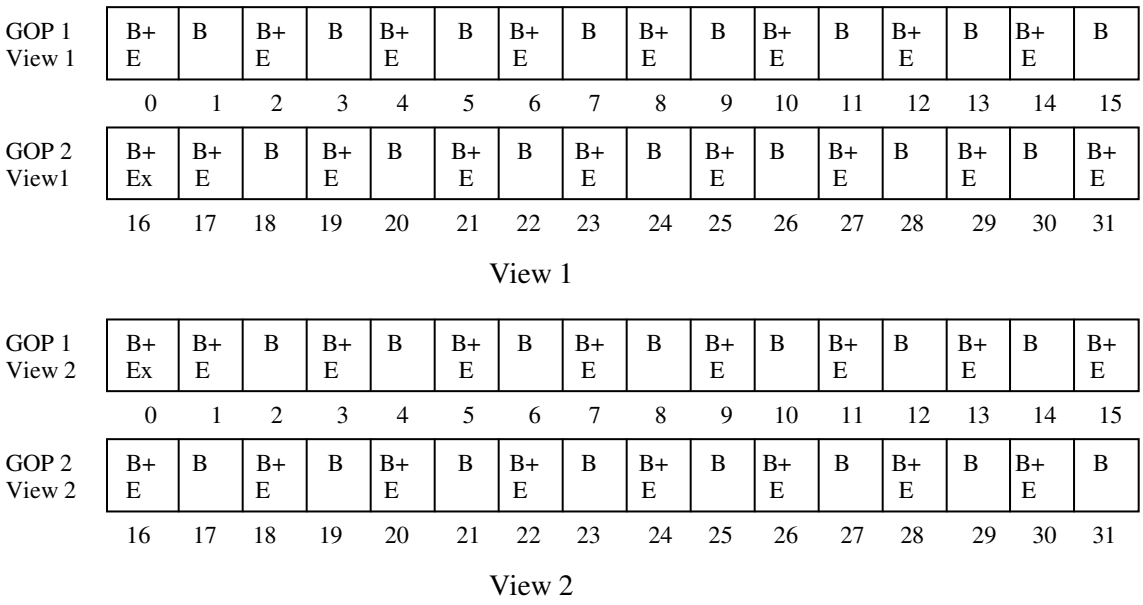


Figure 3.3: Description 1 of stereo video using the General Method 1.

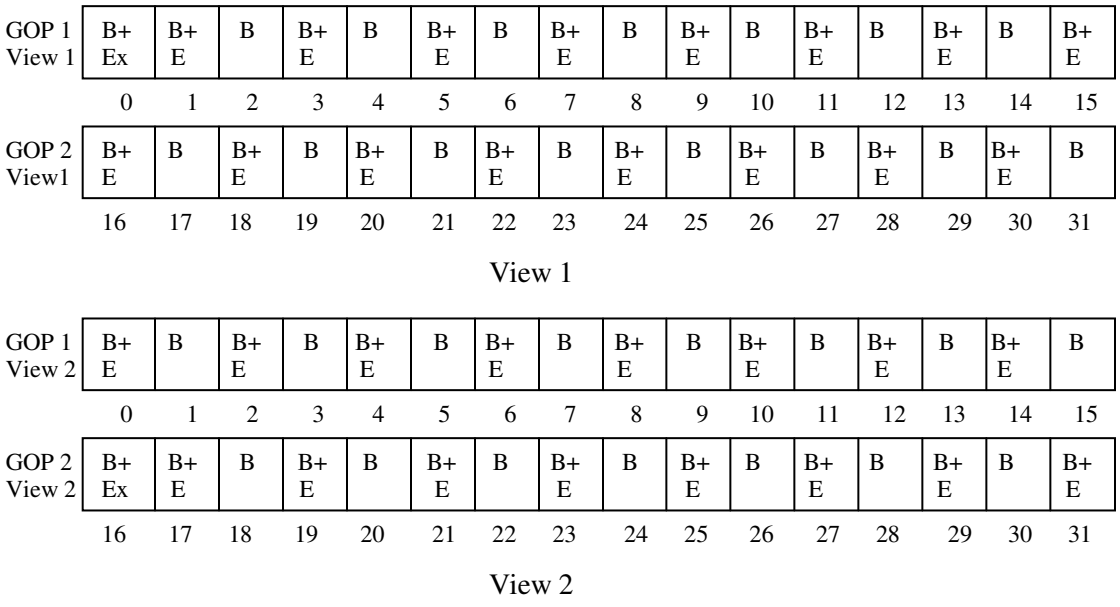


Figure 3.4: Description 2 of General method 1 for Scalable MDC for stereo videos

Secondly, as shown in Figure 3.5 and 3.6, base layer and all MGS NAL units for view 1 and only base layer NAL units for view 2 can be coupled in description 1, and base layer and all MGS NAL units for view 2 and base layer NAL units for view 1 can be coupled in description 2. The extension of these SMDC generation methods from stereo to multi-view video is straightforward.

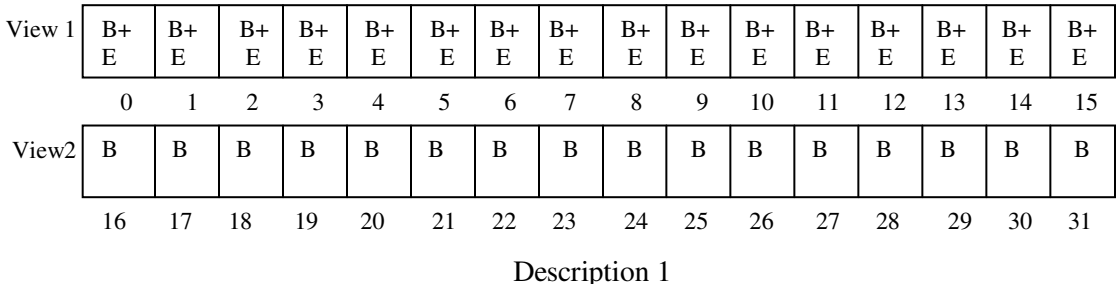


Figure 3.5: Description 1 of General method 2 for Scalable MDC for stereo videos

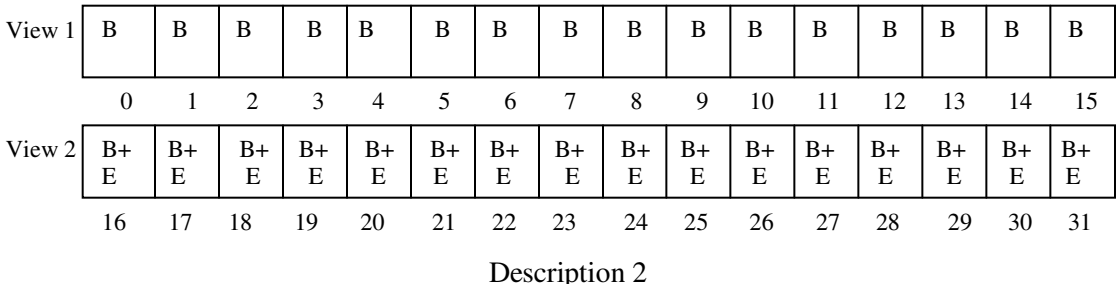


Figure 3.6: Description 2 of General method 2 for Scalable MDC for stereo videos

3.3.2 Definition of Objective Criteria for Stereo Video

Given a stereo description j at the extraction point i of a particular encoding configuration, we define the total distortion of the stereo description $D_j^s(i)$ by

$$D_j^s(i) = \frac{D_j^{v1}(i) + D_j^{v2}(i)}{2} \tag{3.13}$$

where $D_j\_L(i)$ and $D_j\_R(i)$ stand for the distortion of left view and the right view, respectively. Next, we calculate the mean distortion values similar to the monocular case by means of a Monte Carlo Simulation. Then, we define the mean rate-distortion objective of a scalable stereo description as the average of RD values over a set of sample extraction points as follows

$$RD_j\_tot = \frac{\sum_{i=1}^{N_{EP}} \left[\bar{D}_j\_tot(i) + \lambda_i R_j\_tot(i) \right]}{N_{EP}} \tag{3.14}$$

where $R_j\_tot(i)$ is the sum of the right view rate and left view rate.

We compute the mean rate-distortion objective of each stereo description j for N different encoding configurations, denoted by $RD_j\_tot(k)$, k=1,…, N. We finally normalize them to the range [0,1] by

$$NRD_j\_tot(k) = \frac{\min\{RD_j\_tot(1),\ldots,RD_j\_tot(N)\}}{RD_j\_tot(k)} \tag{3.15}$$

We define coverage of a stereo description j of configuration k as the bitrate range of allowed extraction points

$$C_j\_stereo(k) = MaxBR_j\_tot(k) - MinBR_j\_tot(k) \tag{3.16}$$

where $MaxBR_j\_tot(k)$ is the total of the maximum bitrate for view 1 plus view 2, and $MinBR_j\_tot$ is the total of the minimum bitrate for view 1 plus view 2 for description j of configuration k.

Finally, the redundancy of the SMDC for Stereo videos can be defined as the sum of redundancies of the each view, given by,

$$\rho\_stereo = R\_view1 + R\_view2 - (R\_view1^* + R\_view2^*) \tag{3.17}$$

Then, the relative redundancy of configuration k becomes,

$$RR\_stereo(k) = \frac{\rho\_stereo(k)}{(R\_view1^*(k) + R\_view2^*(k))} \tag{3.18}$$

3.3.3 Multiple-Objective Optimization Formulation

The same MOO formulation as in monocular case is applied to stereo case except the number of possible SMDC encoding configurations is higher. The objective functions of

MOO formulation are equations 3.15, 3.16 and 3.18. The same scaling methods as monocular case are used, so the distance formulas remain the same.

3.4 Experimental Results

In our experiments, we used two stereo video pairs, “rena42 - rena43” and “flowerpot0 - flowerpot1.” Both views of the video “rena” is 640 x 480 at 30 frames/sec. and “flowerpot” is 720 x 480 at 25 frames/sec. For the monocular video experiments, the views “rena43” and “flowerpot0” have been used. We also used the video “soccer” which is 720 x 576 at 25 frames/sec for the monocular experiments.

3.4.1 Results for Monocular Videos

There are several encoding configuration parameters, some of which are shown in Table 3.1. Whereas some of these parameters are selected as optimization variables, others have been fixed in our experiments for the sake of simplicity. The fixed parameters are GOP size, enhancement layer QP value, Intra Period, MGS vector type. The variables are the base layer QP and multiple description generation methods. GOP size is set to 16, since decreasing the GOP size to 8 or 4 will increase redundancy. The Intra Period is also set to 16, such that frames at the beginning of each GOP are chosen as I frames. Since we use hierarchical B-pictures with the number of reference frames set equal to 1, drift will be limited to at most within two GOPs if base layer of any frame is lost.

Parameter	Typical Value	Definition
GOPSize	16	GOP Size (at maximum. frame rate)
IntraPeriod	16	
BaseLayerMode	2	Base layer mode (0: AVC w larger DPB, 1:AVC compatible, 2:AVC w subseq SEI)
CgsSnrRefinement	1	SNR refinement as 1: MGS; 0: CGS
EncodeKeyPictures	2	Key pics at T=0 (0:none, 1:MGS, 2:all)
MGSControl	1	ME/MC for non-key pictures in MGS layers (0:std, 1:ME with EL, 2:ME+MC with EL)
MGSKeyPicMotRef	1	motion ref for MGS key pics (0:off, 1:on)
SearchMode	4	Motion search mode (0: BlockSearch, 4: FastSearch)
SearchRange	96	Motion search range (Full Pel)
NumLayers	2	Number of layers
Layer0 QP	Optimization variable	Base Layer Quantization Parameter
Layer1 QP	Parameter fixed for each video	Enhancement Layer Quantization Parameter
MGSVectorType	111111111223	Grouping of DCT coefficients

Table 3.1: JSVM encoding parameters

Table 3.2 and Table 3.3 show Rate-Distortion Objective and redundancy values for different enhancement layer QP and base layer QP values for 2 different videos, “rena42” and soccer_4cif.” For both videos, the RD and redundancy values depend only on QP difference between the base layer and enhancement layer. Therefore, the enhancement layer QP can be fixed and the base layer will become the optimization variable.

	rena42.yuv					
	Normalized Rate-Distortion (NRD)			Relative Redundancy (RR)		
ΔQP	EL QP 26	EL QP 28	EL QP 30	EL QP 26	EL QP 28	EL QP 30
6	1,000	1,000	1,000	0,342	0,341	0,344
8	0,911	0,900	0,897	0,262	0,267	0,264
10	0,824	0,814	0,817	0,202	0,203	0,205
12	0,743	0,742	0,739	0,154	0,157	0,156
14	0,676	0,667	0,671	0,119	0,119	0,124
16	0,603	0,602	0,608	0,091	0,095	0,099

Table 3.2: Rate-Distortion and Redundancy values vs. QP difference

	soccer_4cif.yuv					
	Normalized Rate-Distortion (NRD)			Relative Redundancy (RR)		
ΔQP	EL QP 26	EL QP 28	EL QP30	EL QP 26	EL QP 28	EL QP 30
6	1,0000	1,0000	1,0000	0,414	0,420	0,418
8	0,8974	0,8877	0,8849	0,323	0,325	0,323
10	0,7968	0,7889	0,7881	0,245	0,245	0,245
12	0,7039	0,7031	0,6985	0,184	0,185	0,183
14	0,6263	0,6233	0,6234	0,137	0,136	0,132
16	0,5550	0,5561	0,5607	0,101	0,098	0,096

Table 3.3: Rate-Distortion and Redundancy values vs. QP difference

We encoded input videos with N=20 different encoding configurations, that is, with four different base layer QP and five different MD generation methods. These configurations are shown in Table 3.4 for “rena43” and in Table 3.5 for “flowerpot0”. For “rena43” the enhancement layer QP is fixed to 28 and for “flowerpot0” it is set to 32. Different QP values are used for these two videos because while “rena43” has a fixed background, in “flowerpot0” the camera also moves and the background is not fixed. The PSNR vales for “rena43” are too high because the background is fixed and if we increase the QP value, the visual quality will be not acceptable for the fast moving foreground object. After fixing the enhancement layer QP to a value, that gives acceptable visual quality for the foreground, we increased the base layer QP as much as we can since this

will decrease the redundancy and the redundancy is an important criterion. Packet losses are simulated by N_{MC}=100 Monte Carlo Simulations with 2700 packets, where the packet loss rate is set to p=0.02.

rena43.yuv																
Config. Variables		Config. Objective Values							MOO Results							
									Method1		Method2		Method3		Method4	
EL-B QP	MD Method	NRD_1	NRD_2	NRD_3	C_1	C_2	C_3	RR	Rank	$Dist_1$	Rank	$Dist_2$	Rank	$Dist_3$	Rank	$Dist_4$
28-38	5	0,593	0,610	0,756	324	340	663	0,246	1	0,547	1	0,482	2	0,270	2	0,343
28-38	2	0,642	0,759	0,760	367	379	663	0,339	2	0,551	2	0,502	1	0,256	5	0,481
28-40	2	0,538	0,575	0,665	404	418	729	0,295	3	0,573	3	0,521	3	0,275	3	0,431
28-40	5	0,502	0,528	0,665	355	374	729	0,192	4	0,583	4	0,536	5	0,288	1	0,288
28-36	5	0,730	0,791	0,856	287	300	586	0,318	5	0,595	5	0,565	4	0,275	4	0,465
28-36	2	0,772	0,876	0,861	322	332	586	0,395	6	0,647	6	0,628	6	0,290	6	0,553
28-40	3	0,611	0,634	0,671	466	475	729	0,428	7	0,697	7	0,658	7	0,317	8	0,589
28-38	3	0,706	0,743	0,762	425	432	663	0,465	8	0,719	8	0,686	8	0,323	9	0,611
28-34	5	0,825	0,996	0,990	253	265	517	0,398	9	0,747	9	0,742	9	0,331	7	0,578
28-36	3	0,838	0,832	0,867	372	377	586	0,507	10	0,807	10	0,794	10	0,356	11	0,643
28-34	2	0,893	0,925	0,992	282	292	517	0,463	11	0,814	11	0,811	11	0,357	10	0,629
28-40	1	0,632	0,660	0,677	509	514	729	0,517	12	0,849	12	0,818	12	0,381	12	0,659
28-40	4	0,406	0,429	0,668	509	514	729	0,488	13	0,880	15	0,879	15	0,419	14	0,674
28-38	1	0,706	0,724	0,768	466	469	663	0,554	14	0,887	13	0,859	13	0,397	13	0,673
28-38	4	0,496	0,447	0,754	466	469	663	0,518	15	0,893	14	0,866	17	0,421	15	0,683
28-36	4	0,541	0,544	0,857	410	412	586	0,547	16	0,938	16	0,903	18	0,433	18	0,694
28-34	3	1,000	0,930	0,997	325	331	517	0,559	17	0,954	18	0,953	14	0,415	16	0,684
28-36	1	0,842	0,911	0,872	410	412	586	0,591	18	0,958	17	0,950	16	0,421	17	0,688
28-34	4	0,575	0,624	0,975	360	363	517	0,582	19	1,031	19	1,005	19	0,465	19	0,716
28-34	1	0,951	1,000	1,000	360	363	517	0,637	20	1,091	20	1,091	20	0,476	20	0,718

Table 3.4: SMDC configuration optimization results

flowerpot0.yuv																
Config. Variables		Config. Objective Values							MOO Results							
									Method1		Method2		Method3		Method4	
EL-B QP	MD Method	NRD_1	NRD_2	NRD_3	C_1	C_2	C_3	RR	Rank	$Dist_1$	Rank	$Dist_2$	Rank	$Dist_3$	Rank	$Dist_4$
28-38	2	0,563	0,611	0,698	518	520	956	0,314	1	0,559	1	0,487	1	0,309	4	0,481
28-40	2	0,487	0,506	0,580	581	583	1068	0,264	2	0,573	3	0,510	2	0,318	3	0,411
28-38	5	0,574	0,483	0,689	477	480	956	0,251	3	0,579	2	0,508	3	0,324	2	0,395
28-40	5	0,470	0,489	0,580	533	537	1068	0,192	4	0,589	4	0,532	4	0,328	1	0,328
28-36	2	0,699	0,728	0,823	447	449	829	0,385	5	0,627	5	0,592	5	0,334	6	0,569
28-36	5	0,552	0,603	0,829	413	417	829	0,331	6	0,647	6	0,594	6	0,351	5	0,529
28-40	3	0,483	0,500	0,593	685	683	1068	0,419	7	0,661	8	0,606	8	0,366	8	0,613
28-38	3	0,640	0,585	0,697	613	612	956	0,461	8	0,663	7	0,606	7	0,363	9	0,632
28-34	5	0,856	0,826	0,987	363	367	728	0,415	9	0,716	9	0,708	9	0,369	7	0,612
28-36	3	0,809	0,654	0,832	525	524	829	0,509	10	0,738	10	0,710	11	0,395	11	0,665
28-34	2	0,827	0,810	0,989	392	394	728	0,460	11	0,751	11	0,741	10	0,390	10	0,647
28-34	3	1,000	0,860	0,992	459	459	728	0,567	12	0,850	12	0,847	12	0,445	12	0,703
28-40	1	0,572	0,507	0,596	831	826	1068	0,637	13	0,934	13	0,894	13	0,513	13	0,743
28-38	4	0,539	0,432	0,699	751	747	956	0,655	14	0,968	14	0,932	14	0,536	16	0,756
28-40	4	0,394	0,382	0,551	831	826	1068	0,621	15	0,971	16	0,971	15	0,538	18	0,763
28-38	1	0,519	0,571	0,711	751	747	956	0,675	16	0,980	15	0,939	16	0,540	15	0,755
28-36	1	0,792	0,770	0,842	648	644	829	0,704	17	1,001	18	0,986	17	0,540	14	0,747
28-36	4	0,587	0,512	0,833	648	644	829	0,680	18	1,007	17	0,973	18	0,550	17	0,761
28-34	4	0,568	0,595	0,969	569	566	728	0,712	19	1,087	19	1,060	19	0,585	20	0,778
28-34	1	0,687	1,000	1,000	569	566	728	0,742	20	1,100	20	1,091	20	0,588	19	0,770

Table 3.5: SMDC configuration optimization results

The rate vs. PSNR graphs for first four ranked SMDC encoding configurations are given in Figures 3.7, 3.8, 3.9, and 3.10 for the video “rena42,” and in Figures 3.11, 3.12, 3.13, 3.14 for the video “flowerpot0”. The graphs show the rate and PSNR values for the cases when description 1 or 2 is received, R_1, $PSNR_1$, R_2, $PSNR_2$, and when both descriptions are received, R_3, $PSNR_3$. The PSNR values are mean PSNR values, that is to say the results of 100 Monte Carlo simulations.

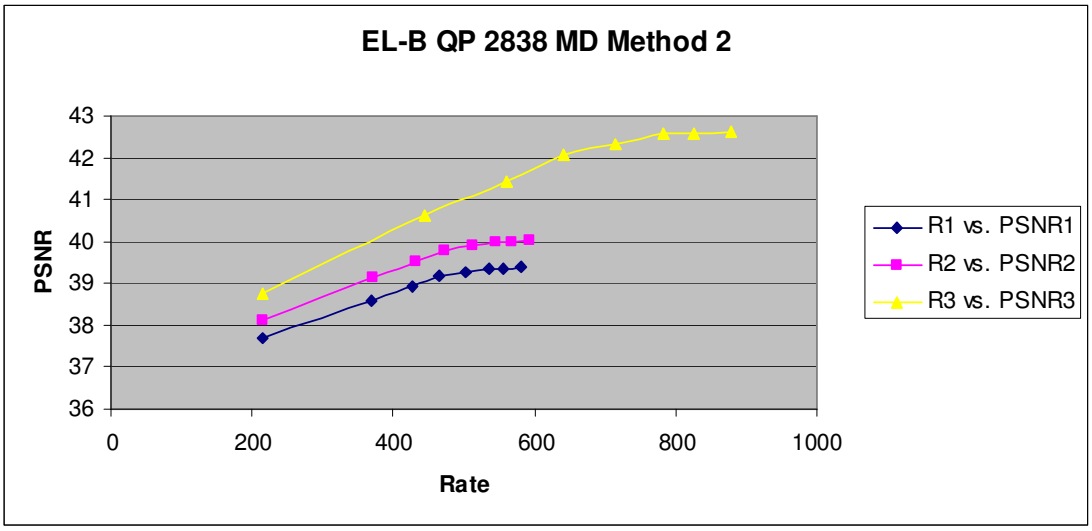


Figure 3.7**:** RD performances of SMDC configuration EL-B 28-38, MD Method 2 for Rena video

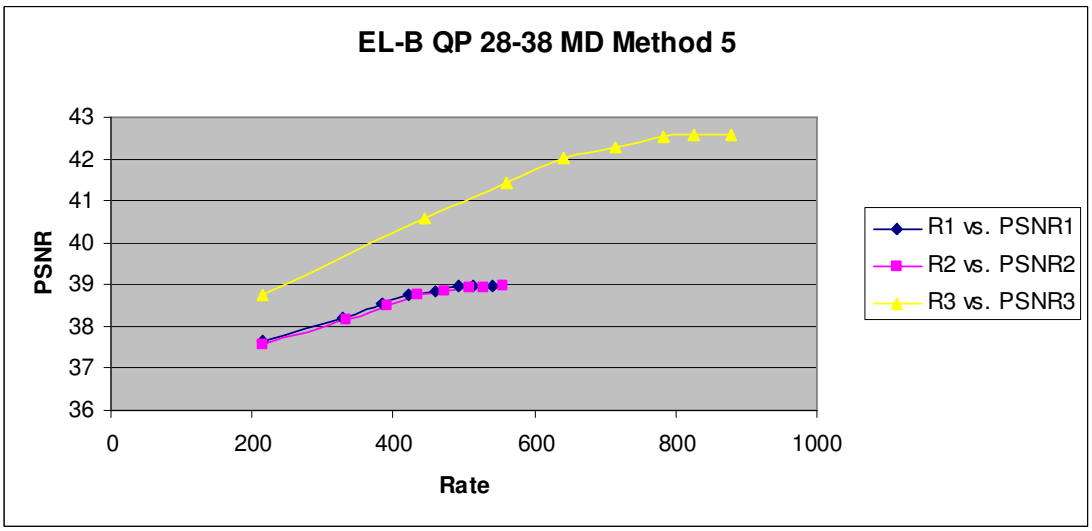


Figure 3.8**:** RD performances of SMDC configuration EL-B 28-38, MD Method 5 for Rena video

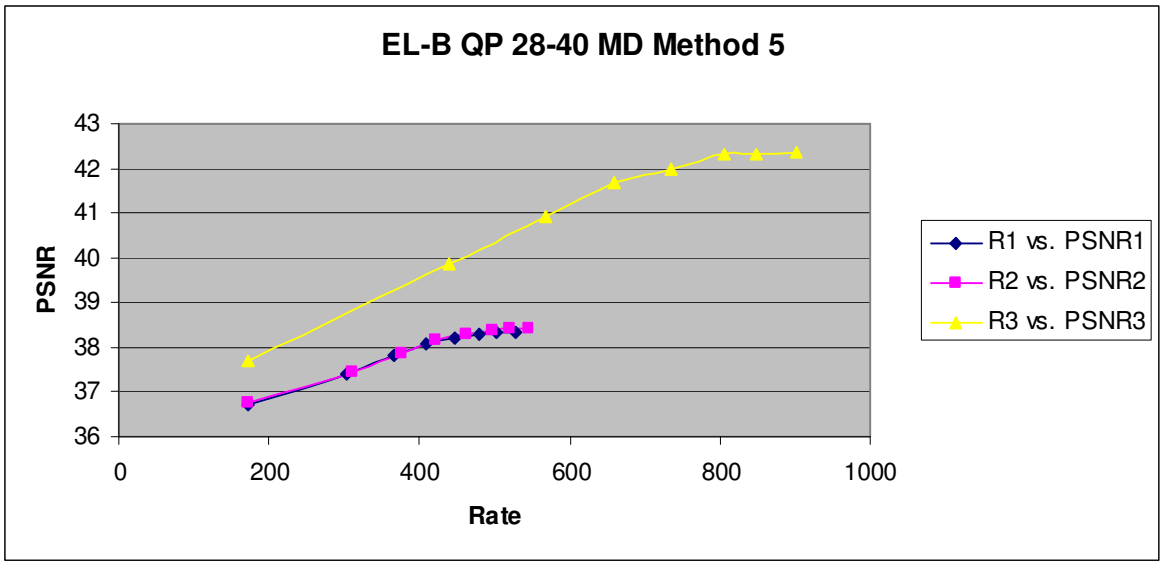


Figure 3.9**:** RD performances of SMDC configuration EL-B 28-40, MD Method 5 for Rena video

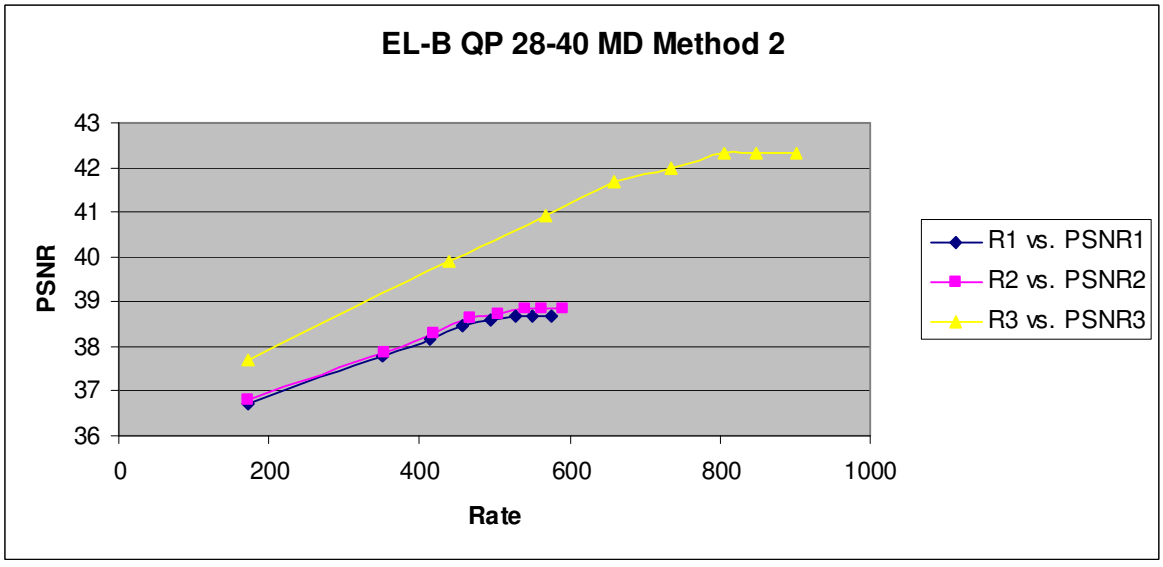


Figure 3.10**:** RD performances of SMDC configuration EL-B 28-40, MD Method 2 for Rena video

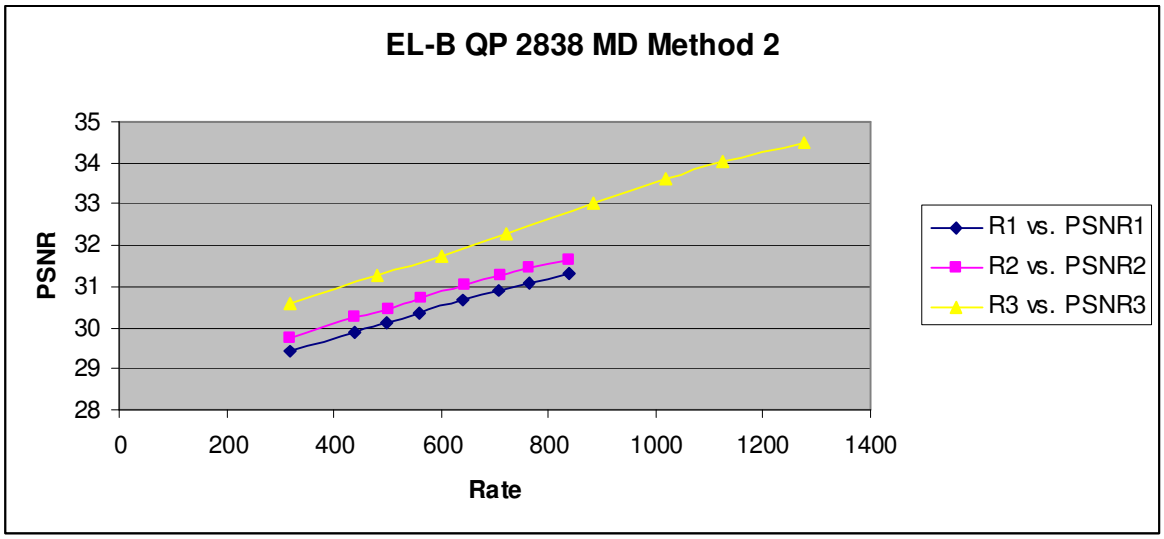


Figure 3.11**:** RD performances of SMDC configuration EL-B 28-38, MD Method 2 for flowerpot video

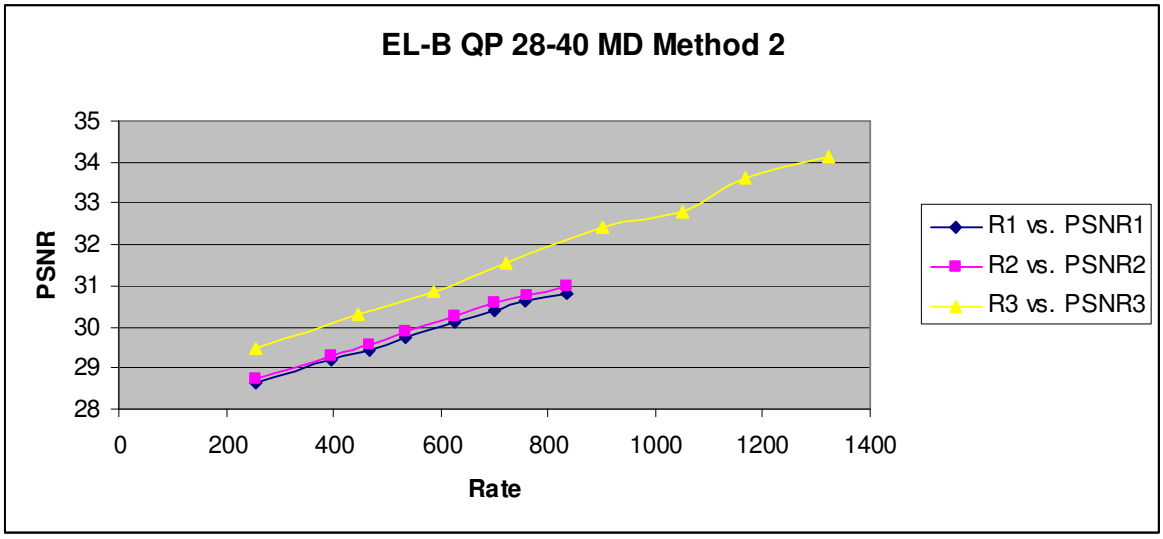


Figure 3.12**:** RD performances of SMDC configuration EL-B 28-40, MD Method 2 for flowerpot video

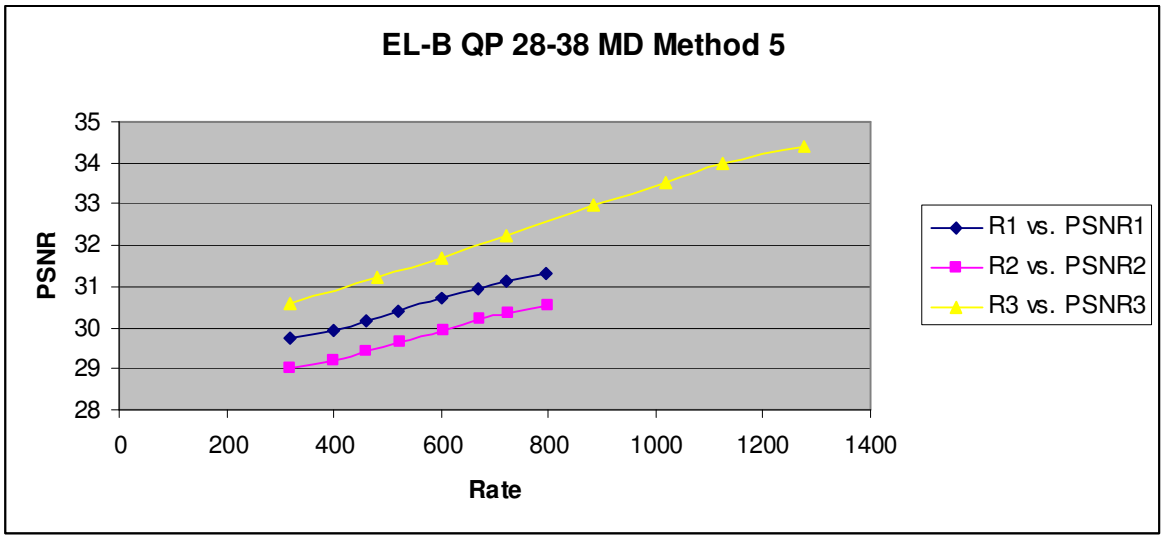


Figure 3.13**:** RD performances of SMDC configuration EL-B 28-38, MD Method 5 for flowerpot video

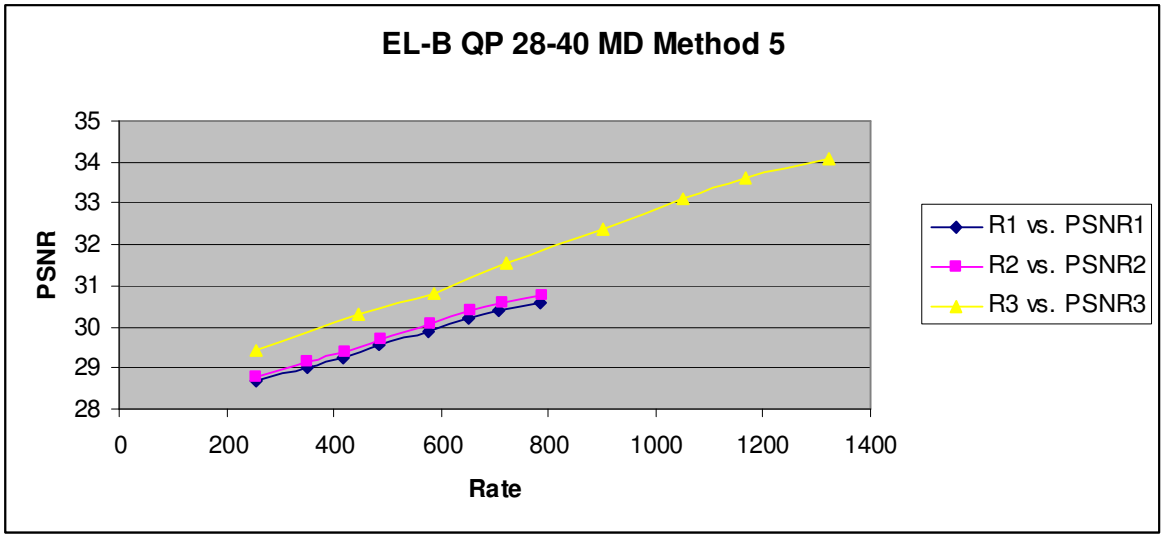


Figure 3.14**:** RD performances of SMDC configuration EL-B 28-40, MD Method 5 for flowerpot video

3.4.2 Results for Stereo Videos

The fixed encoding configuration parameters for the stereo case are set to the same values as in the monocular case. There are N=24 different encoding configurations, four base layer QP value, same as monocular case, and six MD generation methods, five of them same as the monocular case, the general method one, and one more, as explained in Section 3.3.

For calculating mean PSNR values for possible extraction points again a Monte Carlo Simulation with 0.02 packet loss probability is used. There are different strategies to pair extraction points of the two views. While trying to match the TFRC rate, since we simulcast the views, we can first decrease the rate of one view, but fix the rate of the other view, or we can decrease the rate of the both views. While calculating RD values, we considered both of these scenarios to calculate rate and PSNR values for possible extraction points.

Table 3.6 and Table 3.7 show the results of MOO for videos rena and flowerpot, respectively. Additionally, Table 3.8 and Table 3.9 show rate and PSNR values for possible extraction points for both views. R1 of view 1 is coupled with R2 of view 2, and R2 of view 1 is coupled with R1 of view 2 into one description. For example, for 28-40 MD 2 configuration in Table 3.8, the total rate for description 1 is 1678 (688 of view1 + 590 of view 2) and the minimum rate is 368 (196 of view 1 + 172 of view 2). We can combine these six extraction points of each view together. For example, while matching the TFRC rate, we can decrease rate of each view together, view 1 to 520, and view 2 to 466, or we can decrease only the rate of the second view to 420. All the combinations for these extraction points are considered while calculating the rate-distortion objective.

The same scaling methods as in the monocular MOO formulation are used.

rena42.yuv & rena43.yuv																
Config. Variables		Config. Objective Values							MOO Results							
									Method1		Method2		Method3		Method4	
EL-B QP	MD Method	NRD_1	NRD_2	NRD_3	C_1	C_2	C_3	RR	Rank	$Dist_1$	Rank	$Dist_2$	Rank	$Dist_3$	Rank	$Dist_4$
28-38	6	0,640	0,651	0,696	663	839	1502	0,234	1	0,51	1	0,44	2	0,265	3	0,338
28-38	5	0,629	0,640	0,696	749	756	1502	0,235	2	0,516	2	0,445	3	0,265	4	0,341
28-40	2	0,590	0,573	0,627	911	917	1640	0,277	3	0,524	3	0,449	5	0,27	5	0,422
28-38	2	0,696	0,659	0,721	831	838	1502	0,319	4	0,529	4	0,468	1	0,261	8	0,48
28-40	6	0,548	0,535	0,602	729	912	1640	0,183	5	0,546	5	0,479	9	0,291	2	0,291
28-40	5	0,536	0,531	0,604	818	825	1640	0,184	6	0,549	6	0,483	8	0,29	1	0,29
28-36	6	0,782	0,762	0,819	586	753	1339	0,303	7	0,566	7	0,535	6	0,271	7	0,464
28-36	5	0,802	0,754	0,824	667	674	1339	0,304	8	0,566	8	0,537	4	0,268	6	0,464
28-36	2	0,892	0,790	0,856	734	741	1339	0,373	9	0,608	9	0,592	7	0,278	9	0,548
28-40	3	0,628	0,606	0,645	1041	1047	1640	0,406	10	0,651	10	0,594	11	0,314	12	0,592
28-38	3	0,744	0,760	0,759	952	958	1502	0,442	11	0,665	11	0,632	10	0,307	13	0,608
28-34	6	0,936	0,892	0,960	517	674	1191	0,380	12	0,711	12	0,707	13	0,32	11	0,576
28-34	5	0,977	0,929	0,979	595	599	1191	0,381	13	0,711	13	0,71	12	0,316	10	0,575
28-36	3	0,809	0,862	0,856	839	845	1339	0,483	14	0,767	14	0,754	14	0,344	15	0,643
28-34	2	0,887	0,914	0,958	651	655	1191	0,439	15	0,778	15	0,775	15	0,344	14	0,627
28-40	1	0,643	0,631	0,658	1153	1156	1640	0,516	16	0,84	16	0,798	16	0,391	16	0,677
28-38	1	0,874	0,739	0,752	1060	1063	1502	0,551	17	0,867	17	0,847	17	0,392	17	0,681
28-40	4	0,429	0,421	0,549	1153	1156	1640	0,489	18	0,894	19	0,894	21	0,435	19	0,698
28-38	4	0,480	0,501	0,629	1060	1063	1502	0,517	19	0,907	18	0,873	20	0,432	21	0,702
28-34	3	0,975	1,000	1,000	742	745	1191	0,534	20	0,913	21	0,913	18	0,4	18	0,684
28-36	4	0,548	0,590	0,723	940	943	1339	0,545	21	0,947	20	0,909	22	0,438	22	0,708
28-36	1	0,809	0,890	0,877	940	943	1339	0,586	22	0,956	22	0,947	19	0,425	20	0,701
28-34	4	0,631	0,640	0,823	836	837	1191	0,580	23	1,032	23	1,003	23	0,464	23	0,725
28-34	1	1,000	0,978	0,998	836	837	1191	0,631	24	1,084	24	1,084	24	0,475	24	0,727

Table 3.6: Stereo SMDC configuration optimization results

flowerpot0.yuv & flowerpot1.yuv																
Config. Variables		Config. Objective Values							MOO Results							
									Method1		Method2		Method3		Method4	
EL-B QP	MD Method	NRD_1	NRD_2	NRD_3	C_1	C_2	C_3	RR	Rank	$Dist_1$	Rank	$Dist_2$	Rank	$Dist_3$	Rank	$Dist_4$
28-38	2	0,632	0,620	0,681	1024	1024	1884	0,314	1	0,559	1	0,492	1	0,3	6	0,479
28-40	2	0,549	0,555	0,571	1148	1147	2103	0,264	2	0,563	2	0,502	2	0,301	5	0,403
28-40	5	0,523	0,511	0,617	1053	1053	2103	0,191	3	0,576	5	0,52	3	0,311	1	0,311
28-38	6	0,581	0,587	0,648	928	956	1884	0,249	4	0,58	3	0,513	4	0,311	3	0,385
28-38	5	0,618	0,542	0,644	943	943	1884	0,250	5	0,584	4	0,52	5	0,313	4	0,388
28-36	6	0,743	0,699	0,788	804	829	1633	0,329	6	0,61	6	0,571	6	0,321	7	0,511
28-36	5	0,680	0,670	0,768	818	817	1633	0,330	7	0,631	7	0,584	7	0,332	8	0,519
28-36	2	0,765	0,690	0,816	884	884	1633	0,384	8	0,631	8	0,596	8	0,333	9	0,572
28-40	6	0,487	0,449	0,532	1035	1068	2103	0,191	9	0,635	11	0,621	9	0,34	2	0,34
28-40	3	0,542	0,536	0,570	1354	1353	2103	0,423	10	0,665	10	0,617	10	0,358	12	0,613
28-38	3	0,688	0,604	0,678	1212	1211	1884	0,464	11	0,668	9	0,616	12	0,362	13	0,635
28-34	6	0,980	0,950	0,994	709	728	1437	0,412	12	0,702	12	0,701	11	0,361	10	0,608
28-34	5	1,000	0,903	0,985	720	719	1437	0,413	13	0,704	13	0,702	13	0,362	11	0,61
28-36	3	0,905	0,822	0,851	1039	1038	1633	0,511	14	0,714	14	0,702	14	0,381	15	0,66
28-34	2	0,827	0,864	0,919	776	776	1437	0,459	15	0,749	15	0,739	15	0,388	14	0,649
28-34	3	0,957	1,000	1,000	909	908	1437	0,567	16	0,845	16	0,845	16	0,444	16	0,705
28-40	1	0,638	0,594	0,605	1637	1637	2103	0,641	17	0,92	17	0,882	17	0,503	17	0,738
28-40	4	0,454	0,456	0,519	1637	1637	2103	0,625	18	0,974	20	0,972	18	0,53	22	0,758
28-38	1	0,651	0,599	0,674	1481	1481	1884	0,678	19	0,975	18	0,939	19	0,533	19	0,751
28-38	4	0,555	0,506	0,623	1481	1481	1884	0,659	20	0,984	19	0,95	20	0,537	20	0,757
28-36	1	0,935	0,809	0,858	1278	1278	1633	0,708	21	0,994	22	0,986	21	0,539	18	0,747
28-36	4	0,650	0,620	0,746	1278	1278	1633	0,684	22	1,007	21	0,975	22	0,545	21	0,757
28-34	4	0,699	0,691	0,858	1124	1124	1437	0,715	23	1,075	23	1,054	23	0,575	24	0,771
28-34	1	0,829	0,970	0,968	1124	1124	1437	0,744	24	1,092	24	1,088	24	0,585	23	0,768

Table 3.7: Stereo SMDC configuration optimization results

		rena42.yuv						rena43.yuv					
	EP #	R_1	$PSNR_1$	R_2	$PSNR_2$	R_3	$PSNR_3$	R_1	$PSNR_1$	R_2	$PSNR_2$	R_3	$PSNR_3$
28-40 MD 2	1	196	37,00	196	36,88	196	37,75	172	36,72	172	36,81	172	37,68
	2	392	38,12	394	37,94	497	39,95	352	37,76	354	37,85	440	39,88
	3	466	38,50	474	38,35	651	40,90	414	38,17	420	38,29	569	40,92
	4	520	38,84	531	38,71	762	41,81	457	38,46	466	38,61	658	41,70
	5	571	38,97	586	38,86	868	42,18	494	38,57	506	38,73	735	42,00
	6	688	39,09	709	38,98	1108	42,49	576	38,69	590	38,86	901	42,34
28-40 MD 6	1	196	37,00	196	36,74	196	37,75	172	36,72	172	36,83	172	37,68
	2			497	38,28	497	39,95			440	38,37	440	39,88
	3			651	39,16	651	40,90			569	39,34	569	40,92
	4			762	39,99	762	41,81			658	40,07	658	41,70
	5			868	40,33	868	42,18			735	40,35	735	42,00
	6			1108	40,62	1108	42,49			901	40,68	901	42,34
28-40 MD 5	1	196	36,86	196	36,73	196	37,76	172	36,72	172	36,76	172	37,69
	2	343	37,60	350	37,56	497	39,94	303	37,42	310	37,46	440	39,88
	3	417	37,95	429	37,94	651	40,89	365	37,81	376	37,87	569	40,91
	4	471	38,26	487	38,28	762	41,79	409	38,09	422	38,17	658	41,70
	5	522	38,39	542	38,41	868	42,15	446	38,21	462	38,28	735	42,00
	6	639	38,49	665	38,52	1108	42,47	528	38,32	546	38,41	901	42,34
28-38 MD 6	1	243	37,58	243	37,94	243	38,81	215	37,71	215	37,74	215	38,76
	2			506	39,16	506	40,64			445	38,83	445	40,61
	3			645	39,88	645	41,40			559	39,60	559	41,44
	4			747	40,55	747	42,10			642	40,17	642	42,06
	5			849	40,85	849	42,42			716	40,41	716	42,32
	6			1082	41,11	1082	42,70			879	40,68	879	42,61
28-38 MD 5	1	243	37,85	243	37,83	243	38,71	215	37,66	215	37,58	215	38,77
	2	372	38,37	378	38,48	506	40,52	327	38,22	333	38,16	445	40,59
	3	439	38,66	450	38,79	645	41,27	383	38,54	392	38,50	559	41,41
	4	488	38,92	502	39,07	747	41,97	423	38,77	435	38,75	642	42,03
	5	538	39,04	555	39,20	849	42,28	459	38,86	473	38,84	716	42,29
	6	652	39,14	675	39,30	1082	42,56	539	38,96	556	38,95	879	42,58

Table 3.8 Rate and PSNR values for top ranked configurations for Stereo MOO results

		flowerpot0.yuv						Flowerpot1.yuv					
	EP #	R_1	$PSNR_1$	R_2	$PSNR_2$	R_3	$PSNR_3$	R_1	$PSNR_1$	R_2	$PSNR_2$	R_3	$PSNR_3$
28-38 MD2	1	319	29,39	319	29,72	319	30,25	305	27,71	305	29,29	305	30,48
	2	440	29,86	440	30,20	480	30,93	424	28,29	424	29,77	462	31,20
	3	500	30,07	500	30,41	600	31,36	483	28,45	483	29,97	579	31,63
	4	560	30,33	561	30,67	720	31,91	542	28,67	544	30,24	699	32,21
	5	641	30,65	643	30,98	884	32,60	623	28,88	626	30,56	863	32,93
	6	838	31,26	839	31,58	1276	34,03	808	28,97	811	31,10	1232	34,23
28-40 MD2	1	253	28,63	253	28,74	253	29,46	242	28,97	242	28,56	242	29,71
	2	397	29,19	396	29,31	446	30,33	385	29,63	384	29,18	432	30,65
	3	467	29,43	466	29,56	585	30,85	452	29,87	451	29,42	567	31,18
	4	534	29,74	535	29,89	721	31,56	519	30,21	520	29,76	703	31,92
	5	624	30,11	626	30,28	902	32,41	609	30,60	611	30,14	883	32,81
	6	834	30,80	836	30,98	1322	34,13	806	31,23	808	30,77	1277	34,34
28-40 MD5	1	253	28,67	253	28,78	253	29,45	242	28,64	242	28,46	242	29,70
	2	349	29,01	350	29,14	446	30,30	336	28,99	337	28,87	432	30,62
	3	418	29,24	420	29,39	585	30,82	404	29,23	405	29,11	567	31,16
	4	486	29,55	489	29,71	721	31,53	471	29,55	474	29,44	703	31,90
	5	576	29,91	580	30,09	902	32,39	561	29,93	565	29,82	883	32,81
	6	786	30,57	790	30,77	1322	34,10	758	30,54	762	30,43	1277	34,38
28-38 MD6	1	319	29,39	319	29,61	319	30,25	305	27,71	305	28,68	305	30,48
	2			480	30,09	480	30,93			462	29,12	462	31,20
	3			600	30,51	600	31,36			579	29,52	579	31,63
	4			720	31,05	720	31,91			699	30,07	699	32,21
	5			884	31,71	884	32,60			863	30,75	863	32,93
	6			1276	33,06	1276	34,03			1232	31,99	1232	34,23
28-38 MD5	1	319	29,74	319	29,00	319	30,56	305	29,33	305	29,29	305	30,82
	2	399	29,95	400	29,21	480	31,23	383	29,58	384	29,49	462	31,55
	3	459	30,15	461	29,40	600	31,67	441	29,78	443	29,69	579	31,99
	4	518	30,41	522	29,65	720	32,25	500	30,05	504	29,95	699	32,60
	5	600	30,72	604	29,95	884	32,95	582	30,39	586	30,27	863	33,36
	6	796	31,33	800	30,52	1276	34,41	767	30,97	771	30,80	1232	34,75

Table 3.9 Rate and PSNR values for top ranked configurations for Stereo MOO results

3.5 Discussion

For the monocular case, the visual quality of the descriptions changes according to base layer QP and MD generation method. Additionally, odd frames use even frames as reference frames due to hierarchical coding structure. So, when even frames include only the base layer, even if the odd frames include both the base layer and all MGS layers, their visual quality will be limited. Therefore, there is a difference between the visual qualities of consecutive GOPs, causing a flashing effect. However, this difference is not the same for each MD generation method, and each value of base layer. The flashing effect is noticed at the places of the video where there is periodic texture. When the QP difference between the layers is equal to 6, this effect is almost unnoticeable for all MD generation methods. However, when we increase the QP difference, especially, when it is equal to 12, the effect is easily noticed, for the MD generation method 5. Increasing the redundancy while adding MGS layers to the frames at the beginning of each GOP, we decrease this effect. For example, when the QP difference is equal to 10, this effect is almost unnoticeable when the MD generation method 2 is used.

Human eye can perceive 3D videos in high quality, even if one of the videos is in high quality, and the other is low quality. Therefore, this flashing effect disappears in the stereo case for all the MD generation methods and for all the base layer QP values. But, the difference in overall visual quality continues.

In stereo case, method 5 and method 6 have the same redundancy, only the base layer. However, the extraction points for the views differ. For example, if we look at Table 3.8, we see that 28-40 MD 6 configuration only have one extraction point for description 1 for rena42.yuv video. There is no visual quality difference when we watch stereo videos for one of the descriptions generated by these two methods.

The scaling of objective criteria has an effect on the weighting of objective criteria. However, how important an objective criterion is, depends on the situation. For example, if there is no need for rate adaptation, in other words if the channel rate is stable, the coverage objective criteria loses importance. However, if the channel rate is instable and changes drastically, we require a higher range of extraction points. Additionally, if the probability to lose one of the descriptions is high, we require that each description has a high quality, we can accept higher redundancy values. Therefore, choice of scaling method depends on experimental environment. Nonetheless, if we analyze all MOO results tables, we see that even if we change the scaling method, the configuration within the top five ranked and lowest five ranked remain the same, except stereo case Method 4 which gives too much weight on redundancy criteria and ranks the configuration according to their redundancy values.

Chapter 4

CONCLUSION

If we consider only error resilience performance at a specific rate, we can state that ARQ and FEC may perform better than scalable multiple description coding method proposed in chapter 3. Nevertheless, ARQ and FEC do not offer path diversity. By means of MDC we can only receive one description and watch the video with acceptable quality. We can achieve path diversity by ARQ and FEC by sending the same description over parallel independent paths, but in this case the rate of ARQ and FEC will increase as well. Additionally, if we do not use any error correction method, SMDC will perform better. For example, Monte Carlo Simulation results comparing SMDC method with H.264/AVC standard shows that SMDC has a better rate-distortion performance, which increase with packet loss rate increase. Combining two video coding standards scalable video coding and multiple description coding, combines the use of path diversity and rate adaptation for video delivery over DCCP. Previous work, which offers this combination, consists of a FEC packetization concept and of wavelet coding standard, not of a scalable coding approach based on H.264/AVC coding standard.

The main advantage of using Multiple-Objective Optimization to the selection of encoding configuration for both SVC and SMDC is that it offers the consideration of multiple objective criteria at the same time, each depending on different aspects, one depending of coding efficiency, the other on network parameters, etc. Hence, MOO can be considered as analogous to Lagrangian Optimization with multiple Lagrange parameters.

However, in Lagrangian Optimization the weights, the Lagrange parameters, are explicit, whereas in MOO the implicit weights are inserted by scaling the objective criteria to different ranges. Since, different objective functions have different units, one unit change for different objectives is not comparable, and this depends on personal costs. In other words, it is not explicit how much redundancy increase is acceptable, for coverage increase of 5 kbits/sec, this depends on how stable the channel rate is, how much rate adaptation is needed. With different scaling options, the weights on the objective criteria can be adapted.

Chapter 2 MOO results indicates that according to how much spatial and temporal detail a video contains influences the best encoding configuration. That is to say, for different videos, different encoding configurations can be selected the best. Nonetheless, there is no drastic change in the ranking of top five configurations from one video to another.

The effect of different scaling options on the results of MOO is studied in Chapter 3. the results show that even if we change the scaling option, for example scaling by normalization by fixing the maximum value to 1, or scaling with stretching to the range [0,1] by fixing maximum value to 1 and minimum value to 0 , the top five configurations remains the same for both videos in monocular case. For stereo case, only one method, which gives too much importance on the redundancy objective criteria ranking the configurations by their redundancy values, includes encoding configuration which is within top ten for other scaling methods in the top five configurations.

Even if when we decrease redundancy, the visual quality of the descriptions decrease, all the methods offers acceptable visual quality, especially in stereo case. Since human eye can perceive 3D videos in high quality, even if one of the videos is in high quality, and the other is low quality, we arranged the descriptions in stereo case in a way that for each frame if the quality of view 1 is low, the quality of view 2 is high, and vice versa.

Therefore, the visual artifacts in low quality videos are eliminated by the video in high quality.

Moreover, how much visual quality is gained in stereo videos by changing the amount of visual quality difference between views is another research question which is part of the future work. Therefore, by this work, we can answer the question of how the rate adaptation should be performed when we have independently scalable bitstreams for the two views.

BIBLIOGRAPHY

[1] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, "Scalable Video Coding Working Draft 1," Doc. JVT-N020, Jan. 2005.

[2] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, "Joint Scalable Video Model JSVM-6.7," Doc. JVT-Q202, Oct. 2005.

[3] H. Schwarz, D. Marpe, T. Schierl, and T. Wiegand, "Combined scalability support for the scalable extension of H.264/AVC," *IEEE Int. Conf. Multimedia & Expo (ICME)*, Amsterdam, The Netherlands, July 2005.

[4] I. Amonou, N. Cammas, S. Kervadec, and S. Pateux, "Optimized rate-distortion extraction with quality layers," *Proc. IEEE Int. Conf. on Image Processing*, Atlanta, Georgia, pp. 173-176, Oct. 2006.

[5] H. Schwarz and T. Wiegand, "Preliminary results for an r-d optimized multi-loop SVC encoder," JVT Doc. JVT-T080, Klagenfurt, July 2006.

[6] T. Weigand, H. Schwarz, A. Joch, F. Kossentini, and G. Sullivan, "Rate_constrained coder control and comparison of coding standards," *IEEE Trans. on Circuits and Systems for Video Technology*, July 2003.

[7] J. Lin, "Multiple objective problems: Pareto-optimal solutions by method of proper equality constraints," *IEEE Trans. Automatic Control*, vol. 21, pp. 641–650, 1976.

[8] E. Kohler, M. Handley, and S. Floyd, "Datagram congestion control protocol (DCCP)." Internet Engineering Task Force, RFC 4340, March 2006.

[9] M. Handley, S. Floyd, J. Padhye, and J. Widmer, "TCP friendly rate control (TFRC): Protocol specification," IETF, RFC 3448, January 2003.

[10] 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 delivery*, vol. 17, no. 9, September 2007.

[11] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, "Joint scalable video model (JSVM) text," Doc JVT-X202, June 2007.

[12] H. Kirchhoffer, D. Marpe, H. Schwarz, and T. Wiegand, "A low-complexity approach for increasing the granularity of packet-based fidelity scalability in scalable video coding," *Picture Coding Symposium*, Lisboa, Portugal, Nov. 2007.

[13] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Mag.,* vol. 18, pp. 74-93, Sep. 2001.

[14] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proc. of IEEE*, vol. 93, no. 1, January 2005.

[15] V. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[16] V. Vaishampayan and J. Domaszewicz, "Design of entropy-constrained multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 40, no. 1, pp. 245–250, January 1994.

[17] V. Vaishampayan and J.-C. Batllo, "Asymptotic analysis of multiple description quantizers," *IEEE Trans. Inform. Theory*, vol. 44, no. 1, pp. 278–284, January 1998.

[18] H. Jafarkhani and V. Tarokh, "Multiple description trellis coded quantization," *Proc. IEEE Int. Conf. Image Processing (ICIP98),* vol. 1, pp. 669–673, October 1998.

[19] V. Vaishampayan, N. Sloane, and S. Servetto, "Multiple description vector quantization with lattice codebooks: design and analysis," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 1718–1734, July 2001

[20] V. Vaishampayan, "Application of multiple description codes to image and video transmission over lossy networks," *Proc. 7th Int. Workshop Packet Video*, Brisbane, Australia, pp. 55–60, March 1996.

[21] S. D. Servetto, K. Ramchadran, V. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *Proc. IEEE Int. Conf. Image Processing,* Chicago, IL, 1998.

[22] ——, "Multiple-description wavelet based image coding," *IEEE Trans. Image Processing,* vol. 9, no. 5, pp. 813–826, May 2000.

[23] S. Diggavi, N. Sloane, and V. Vaishampayan, "Asymmetric multiple description lattice vector quantizers," *IEEE Trans. Inform. Theory*, vol. 48, no. 1, pp. 174–191, January 2002.

[24] V. Goyal, J. Kelner, and J. Kovacevic, "Multiple description vector quantization with a coarse lattice," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 781–788, January 2002.

[25] J. Kelner, V. Goyal, and J. Kovacevic, "Multiple description lattice vector quantization: variations and extension," *Proc. IEEE Data Compression Conf.*, Snowbird, UT, pp. 480–489, Mar. 2000.

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

[27] M. Orchard, Y.Wang, V. Vaishampayan, and A. Reibman, "Redundancy rate distortion analysis of multiple description image coding using pairwise correlating transforms," *Proc. IEEE Int. Conf. Image Processing*, Santa Barbara, CA, pp. 608–611, Oct. 1997.

[28] Y. Wang and S. Lin, "Error-resilient coding using multiple description motion compensation," *Proc. IEEE Int. Workshop Multimedia Signal Processing (MMSP01)*, pp. 441–446, Oct. 2001.

[29] Y. Wang, M. Orchard, and A. Reibman, "Optimal pairwise correlating transform for multiple description coding," *Proc. Int. Conf. Image Processing*, vol. 1 , pp. 679–683, Oct. 1998.

[30] Y. Wang, A. Reibman, M. Orchard, and H. Jafarkhani, "An improvement to multiple description transform coding," *IEEE Trans. Image Processing*, vol. 50, no. 11, pp. 2843–2854, 2002.

[31] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri, "Multiple description coding for video using motion-compensated prediction," *Proc. IEEE Int. Conf. Image Processing (ICIP99)*, vol. 3, pp. 837–841, Oct. 1999.

[32] ——, "Multiple description coding for video using motion-compensated temporal prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 193–204, 2002.

[33] X. Tang and A. Zakhor, "Matching pursuits multiple description coding for wireless video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, 566–575, 2002.

[34] Y. Wang and S. Lin, "Error resilient video coding using multiple description motion compensation," *IEEE Trans. Circuits, Syst., Video Technol.*, vol. 12, no. 6, pp. 438–452, Jun. 2002.

[35] K. R. Matty and L. P. Kondi, "Balanced Multiple Description Video Coding Using Optimal Partitioning of the DCT Coefficients," *IEEE Trans. On Circ. And Syst. For Video Tech.,* vol. 15, no. 7, July 2005.

[36] D. Comás, R. Singh, A. Ortega, and F. Marqués, "Unbalanced multiple description video coding based on a rate-distortion optimization," *EURASIP J. Appl. Signal Process.*, vol. 2003, no. 1, pp. 81–90, Jan. 2003.

[37] M. van der Schaar and D. S. Turaga, "Multiple description scalable coding using wavelet-based motion compensated temporal filtering," *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, Sep. 2003.

[38] E. Akyol, A. M. Tekalp, M. R. Civanlar, “A flexible multiple description coding framework for adaptive peer-to-peer video streaming,” *IEEE Journal of Selected Topics in Signal Process.*, vol. 1, no. 2, August 2007.

[39] A. Norkin, M. O. Bici, A. Aksay, C. Bilen, A. Gotchev, G. Bozdagi Akar, K. Egiazarian, and J. Astola: "Multiple description coding and its relevance to 3DTV". In Haldun M. Ozaktas, and Levent Onural (Eds.), *Three-Dimensional Television: Capture, Transmission, and Display* (ch. 11), Springer Verlag, 2007

[40] A. Norkin, A. Aksay, C. Bilen, G. B. Akar, A. Gotchev, and J. Astola, “Schemes for multiple description coding of stereoscopic video,” *MRCS 2006, LNCS 4105*, Springer, pp. 730-737, Sept. 2006.

[41] L. P. Kondi, “A rate-distortion optimal hybrid scalable/multiple-description video codec,” *Proc. IEEE Accouss., Speech, and Signal Process. (ICASSP 2004),* vol. 3, pp. 269-272, May 2004.

[42] R. Puri and K. Ramchandran, “Multiple description source coding through forward error correction codes,” *in Proc. 33rd Asilomar Conf. Signals, System Comp.*, vol. 1, pp. 342-346, 1999.

[43] P. A. Chou, H. J. Wang, and V. N. Padmanabhan, “Layered multiple description coding,” *Packet Video Conf.*, Nantes, France, 2003

[44] J. R. Taal, A. Pouwelse, and R. L. Lagendijk, “Scalable multiple description coding for video distribution in P2P networks,” *24th Picture Coding Symposium*, San Francisco, USA, December 2004.

[45] V. Stankovic, R. Hamzaoui, and Z. Xiong, “Robust layered multiple description coding of scalable media data for multicast,” IEEE Signal Proc. Letters, vol. 12, no. 2, Feb. 2005.

VITA

T. BERKİN ABANOZ was born in Adapazari, Turkey on January 2, 1982. She received her B.Sc. degrees in Electrical Engineering from Koç University, Istanbul, in 2006. From September 2006 to September 2008, she worked as a teaching and research assistant in Koç University, Turkey where she completed her M.S. degree. She has published chapters about on her research for the following conferences SIU2007 (Eskişehir, Turkey), ICIP2007 (San Antonio, USA).