Transrating of Coded Video Signals via Optimized Requantization

Michael Lavrentiev

TRANSRATING OF CODED VIDEO SIGNALS VIA OPTIMIZED REQUANTIZATION

RESEARCH THESIS

Submitted in Partial Fulfillment of the Requirements
For the Degree of Master of Science
in Electrical Engineering

Michael Lavrentiev

SUBMITTED TO THE SENATE OF THE TECHNION — ISRAEL INSTITUTE OF TECHNOLOGY

IYAR, 5764 HAIFA APRIL, 2004

The Research Thesis was done under the supervision of Prof. David Malah in the department of Electrical Engineering.

It was supported in part by STRIMM consortium under the MAGNET program of the Ministry of Trade and Industry via the Samuel Neaman Institute.

The generous financial help of the Technion is gratefully acknowledged.

I would like to convey my deepest gratitude to Prof. David Malah for his extremely devoted guidance and support throughout the research.

I would like to thank all the staff of Signal and Image Processing Lab and especially to Nimrod Peleg, Itai Segall, Ziv Rabizanovich and Arik Yavilevich for their help in developing video analysis tool. I am also thankfull to Optibase and particularly to Ilan Daniel, for their help in carrying out the perceptual evaluation tests on their Perceptual Quality Assesment system.

Contents

List of Tables

List of Figures

Abstract

Multimedia content provides rich information to consumers, but also poses challenging problems of management, delivery, access, and retrieval because of data volume and complexity. Digital video needs to be compressed for the purpose of efficient storage and transmission. The common solution is to produce or retrieve from storage a single high quality bitstream, and to match it to each end-user bandwidth constraints by transrating.

The goal of transrating is to reduce the bit-rate of the encoded stream, while preserving the highest possible quality of the rate-reduced video. Many transrating techniques have been developed recently. The naive solution of simple cascading of decoder and encoder is put aside because of its high computational complexity, which is mainly due to the need to re-estimate motion parameters, and quality degradation, caused by imprecise DCT/IDCT matching, additional quantization errors and improper motion estimation in the second generation encoder. Motion estimation errors are caused by working on lossy compressed video instead of the original source. Most of the transrating schemes are operating in the DCT domain and utilize the decisions made by the initial encoder to improve the output video quality. Lagrangian optimization of quantization step-size of each Macro-block has provided the best quality in terms of PSNR over previously developed requantization schemes.

In this research work, different MPEG-2 transrating approaches, based on requantization, are implemented and tested. Lagrangian optimization of the quantization step-size of each Macro-block is compared with a "simple" complexity-based transrating scheme; with cascaded decoding and encoding, and with the direct encoding of the original video sequence at the desired final bit-rate, using a standard TM5 encoder. To reduce requantization errors, MSE and MAP requantization decision approaches are added to the previously developed algorithms.

A novel extension of the Lagrangian optimization method, which optimally modifies quantized DCT indices, is developed and compared to existing methods. The proposed method outperforms all currently known requantization-based transrating approaches. To reduce the complexity of the proposed algorithm, we provide a low complexity trellis-based optimization scheme, and discuss other complexity reduction means as well.

Finally, we propose a simple approach for taking into consideration Human Visual System properties in the transrating process. The input pictures are segmented into areas of textures, smooth areas and boundaries, and the distortion of each block is weighted according to its type. This way the Lagrangian optimization allocates bits to the different segments according to their perceptual importance. Perceptual observations, as well as measurements made by Tektronix's Perceptual Quality Assessment (PQA) tool, show an improvement of output video quality by the proposed schemes.

List of Abbreviations

MSE	Mean-Squared-Error
SE	Squared Error
PSNR	Peak Signal to Noise Ratio
MAP	Maximum A-Posteriori
ML	Maximum Likelihood
MPEG	Moving Picture Expert Group
MPEG-2	video standard developed by MPEG
GOP	Group Of Pictures
MB	Macro-Block
MV	Motion Vector
DCT	Discrete Cosine Transform
IDCT	Inverse Discrete Cosine Transform
FT	Fourier Transform
WT	Wavelet Transform
VLC	Variable Length Coding
VLD	Variable Length Decoding
CBR	Constant Bit Rate
VBR	Variable Bit Rate
HVS	Human Visual System
VO	Video Object
RAG	Region Adjacency Graph
MEFC	Maximum Entropy Fuzzy Clustering
NASM	Noise Adaptive Soft-switching Median

List of Symbols

R	Picture bit-rate
D	Distortion caused to picture
q_k	Requantization step-size of k-th MB
d_k	Distortion caused to k-th MB
r_k	Bits needed to encode k-th MB
v_k	Vector of quantized zig-zag ordered DCT coefficients of k-th MB
J_k	Lagrangian cost function of k-th MB
λ	Lagrangian multiplier
C_k^{ni}	i-th inverse quantized coefficient of n-th block in k-th MB
X_k^{ni}, Y_k^{ni}	Accumulated error in reference frames for C_k^{ni}
$J\_min(i)$	The optimal path cost till i-th stage in trellis
$D_0(i)$	The distortion caused by skipping i-th AC coefficient in trellis
$D(v,i)$	The distortion introduced by choosing v to be i-th quantized coefficient in trellis
run	Number of zeros before current stage in trellis
$R(run,v)$	Bits needed to encode (run,v) pair using VLC
$J(v,i)$	The cost of optimal path till i-th coefficient ending with v
$J_{i,j}$	Optimal j-th block cost function for i-th choice of requantization step-size
q_{opt}	Requantization step-size that provides minimum cost function
$J^{opt}(v)$	Part of $J(v,i)$ influenced by particular v choice
J^{opt}	Part of $J(v,i)$ that was already calculated on previous stages
$Q_{1/2}$	Initial/secondary quantization step-sizes (also used for quantization procedure in block diagrams)
$Q_{1/2}^{-1}$	Initial/secondary inverse quantization procedure
σ_i^2	Estimated variance of i-th coefficient
X	Frame or MB complexity
T	Target bit-rate allocation
S	Bit-rate generated by encoder
$B_{s/n}$	Output buffer size/fullness

Chapter 1

Introduction

Multimedia telecommunication services use a great deal of video material, compressed in different formats for storage and transmission. As of today, MPEG-2 is the main standard for video coding. The flexibility of this standard enables its use in a large variety of applications, including video on demand, digital TV, distance learning, and many others. In coding of video data for transmission, the channel characteristics should be provided to the encoder. Yet, the network today consists of a wide set of heterogeneous interconnected networks. So, transmission parameters, such as supported video formats and available bandwidth, may vary greatly in time and from one end-point to another. This is especially pronounced when the same bit- stream is distributed to several decoders. One of the most common solutions is to produce a single high quality encoded bit stream, which can be either produced by an on-line encoder or retrieved from storage media, and to adapt it to current demands by transrating.

Transcoding, in general, is the process of converting a compressed video format into another compressed format at a possibly different rate. Since our research aims at transrating to the same video format, with particular emphasis on MPEG-2, we limit our discussion here to this case, i.e., to what is known as *transrating*.

1.1 Transrating goals

Transrating to the same video format is usually done in order to reduce the bit-rate of the encoded stream, while preserving the highest possible quality of the rate-reduced video. This goal can be set as to achieve the same subjective quality in the entire frame or mostly in regions that can be defined as more informative.

If we are talking about simultaneous transrating of several or many channels that have to be transmitted over the same Constant Bit Rate (CBR) channel, the goal can be set as to increase the mean quality of every sub-channel by taking advantage of non-constant sub-channel usage by other programs [1].

1.2 Transrating approaches

For a specific bit-stream, bit-rate reduction can be achieved by:

1. Frame-rate reduction [2]. Video is usually encoded using temporal prediction techniques, so that not all types of frames can have their rates reduced without introducing additional errors in other frames. The problems of prediction re-estimation and total error minimization over all frames have to be taken into account.However, if talking about B-frames dropping only, no motion re-estimation problem appear.

2. Requantization (increasing quantization step) or/and by discarding Discrete Cosine Transform (DCT) coefficients [3, 4, 5, 6, 7, 8, 9, 10]. This issue will be

discussed in detail in the sequel.

3. Spatial resolution reduction using frame size rescaling (via decimation) [11, 12, 13]. Spatial resolution reduction can be applied in the DCT domain. Still, prediction re-estimation and error propagation are to be faced.

4. Some other means like picture cropping [14]. If one can define the importance of each part of the image, picture cropping that leaves the most important parts with total bit-rate under a given constraint can be provided. However, these methods need side information about different parts of the video sequence and their relative importance.

In this work we consider only methods of requantization or elimination of DCT coefficients without frame dropping, cropping or re-scaling. These methods can be applied without any additional information about picture content and may be done without prediction re-estimation, which is the most computational intensive operation.

1.2.1 DCT coefficients modification

Several transrating methods were recently discussed in the literature [3, 8, 15, 16, 17]. The straightforward approach is to fully decode the sequence, and to encode it by the same encoder but with more severe constraints. However, this approach is computational expensive, and also introduces requantization errors to the full extent. This is because re-encoding depends on the initial encoding, but does not take it into account in any way. Alternatively, using decisions made in the initial encoding process may help in minimizing the computational load, as well in error reduction. As discussed by [8],[15] transrating approaches can be coarsely classified by the methods used for building the transrating architectures. At the upper level they can be divided as follows:

1. Re-encoding using the original motion vectors and coding decisions such as Macro Block (MB) types, quantization weighting matrix, Group Of Pictures (GOP) and slice (group of MBs) structures, motion vectors, etc. [3], [8].

2. Re-encoding using the original motion vectors information but new coding decisions [8]. For example, in B-pictures the prediction type can be changed from bi-directional to forward or backward only, with a proper update of the motion vectors, but the information about the original prediction area can help in reducing the motion estimation complexity.

Bit-rate reduction in each of the above classes is achieved by one of the following means:

1. Discarding high frequency DCT coefficients [8, 16, 17]. In this method, decoding up to the point (inclusive) of inverse Variable Length Coding (VLC) is performed. At this point the transcoder has exact information about MB bit allocation, and can easily calculate the bit-rate produced after it changes the bit-stream. In the case of discarding of the last coefficients in zig-zag scan, the transcoder can calculate the gain in bit-rate reduction by simply reducing the length of the corresponding VLC words from the input MB bit-rate occupation. In the case of dropping non-zero coefficients from the middle of the zig-zag scan, the VLC lookup tables need to be used to find the change in bit-rate. The rate-distortion optimization problem can be solved as described in the sequel in section 4.2. One more aspect of this method is that the introduced distortion is directly obtained from the discarded coefficients and hence simplifies the

calculations.

2. Requantization by just increasing the quantization step [8]. This is actually a more complex scheme than the previous one. After inverse VLC, the DCT coefficients need to be recovered, using inverse quantization. Bit-rate reduction is achieved by quantizing the coefficients with a larger quantizer step. The assumption is that this will produce an error that is more equally distributed among all DCT coefficients than by discarding some of them. In such a case there is no easy way to predict the introduced error and the resulting bit-rate.

1.2.2 Open-loop vs. closed-loop transrating

When talking about transrating complexity and robustness, open-loop and closed-loop transrating have to be considered. Open-loop transrating refers to transrating of the bit-stream on a frame-by-frame basis, without taking into account the changes done to previously transcoded frames. It is the fastest approach, but it leads to a continuously decreasing quality in each Group Of Pictures (GOP), because in MPEG, as well as in many others video encoding techniques, frame prediction is used, and only frame residuals with prediction parameters are actually sent. As an alternative, closed-loop transrating was proposed in [3], [9]. The idea is to add the information removed by the transrating process in a given frame to the data of the next dependent frames. Alternatively, it can be seen as compensating the error introduced during transrating. The reduction in error propagation is of utmost importance because differential encoding is used in most video encoding systems due to its efficiency. However, there is the problem that the error propagation depends on motion prediction. After intensive research in this area, it was found that if the only change one wants to introduce is in the quantization step, all the computations for closed-loop error compensation can also be done in the DCT domain [10],[18]. Of course, this bring up a whole new set of problems: How to produce motion compensation in the DCT domain [11]; what to do with skipped MBs; what additional changes in bit stream structure have to be done if a change in the GOP structure is necessary (for example, if the rate control provides no bits for the transmission of particular frame). But the most important one is how to obtain the best possible quality at the required output bit-rate.

1.2.3 HVS-based Adaptive Transrating

The quantification of the introduced distortion is typically done via the Mean-Squared-Error (MSE). The problem with the MSE metric is that it is not always a good indicator of picture quality. There is a general agreement that the Human Visual System (HVS) properties have to be taken into account while encoding/transrating of video sequences. The idea behind adaptive requantization is that distortion in a macroblock may be masked in proportion to macroblock activity. Current encoders apply adaptive quantization through video sequence analysis accounting for the HVS characteristics. A standard simulation model known as TM ("Video Codec Test Model") is described in [7]. In this model the quantization steps are obtained by multiplying a base quantization step by a weighting factor determined during the encoding process. Since, according to [7], MB activities cannot be obtained from the bit-stream, unless it is fully decoded to recover the picture, it is necessary to estimate appropriate requantization parameters from the coded information.

1.2.4 Rate control

The main task of the transcoder is to achieve a certain bit-rate reduction. So, one of the most important components of adaptive quantization schemes is a model for the number of bits needed to code a macroblock for different values of the quantization step. Such a model reduces the need to check all the possible quantization steps, hence decreases the computational complexity of the transcoder.

The bit allocation and buffer control processes at the encoder produce base quantization steps for each of the GOP, picture, slice and macroblock layers. Local quantization step on slice or MB level and average quantization step over the picture, obtained from coded data, well reflect the local activity and average activity, respectively [1]. Methods proposed in [1],[19] are based on the idea of providing a ratio between quantization steps at the different layers that are as close as possible to those in the originally encoded stream. The idea of adaptive quantization is also used in [20] to develop transrating in the DCT domain, based on analyzing a range of macroblock coefficients values.

It is also a task of transcoding rate control to define the bit-allocation for every frame in the GOP. Some algorithm also try to preserve the ratio of bit-allocations provided by encoding procedure [21]. Other schemes are using the picture complexity, which is defined as the product of the average quantizer step-size and the number of bits generated, divided by some empirical constant [22, 23, 24, 25]. They provide bit allocation according to frame complexities for the frames of the same type. The results of those schemes have to be further changed to withstand the virtual buffer constraints. In [26] an algorithm based on tracking virtual buffer fullness is proposed. Those methods use the complexity instead of the variance of the block used by standard MPEG-2 TM5 rate control. The above frame-level rate-control schemes will be described in details in section 2.2.

1.2.5 Relation between reconstruction error and bit-rate

According to [16], the DCT coefficients of difference frames are assumed to have a Laplacian pdf (eq. 3.3). It is shown in [3] that if one assumes that the encoding bit-rate is close to the entropy H of the quantized coefficients, then the bit-rate obtained after requantization, by applying de-quantization (with quantization step q_1) and quantization with a new quantization step, q_2, has a step-like decreasing behavior as function of q_2/q_1. At the same time the Peak Signal to Noise Ratio (PSNR) for a given "entropy step" is varying over the corresponding range of q_2/q_1, with a more pronounced variation for small ratio values. Similar results were reported by [27] for a Gaussian pdf. The work in [27] also provides the optimal requantization steps as function of the initial quantization steps that achieve in simulations the best PSNR for a given bit-rate reduction factor. In [27] no comparison with theoretical results is provided. Those results were further extended by [23] for complexity reduction of Lagrangian optimization in open-loop transrating. In [28] it is proposed to disable requantization in the range up to twice the original quantization step, because it was shown there that in this range there is the biggest degradation in PSNR, per bit. For larger quantizer step sizes, the optimization problem was solved using a Lagrangian optimization, as presented in section 4.2.

1.3 Organization of the Thesis

This thesis is organized as follows:

Chapter 2 covers existing transrating architectures, GOP-level bit-rate control schemes, and also presents a simple complexity-based transcoding scheme that we are using as reference to compare the performance of more complex transrating approaches.

In chapter 3, MAP and MSE transrating methods based of DCT coefficients statistical properties are presented. Since in transrating, only quantized coefficients information is available, there is a need to estimate the original distribution parameters. The most commonly used PDF model is the Laplacian distribution model. However, a number of works show that it is not the best one in many cases. We propose to estimate the PDF by linear interpolation of the quantized coefficients distribution. This solution can utilize all the MBs in the frame, even though they are quantized by different quantization step-sizes, and is found to provide better results for MAP transrating and in most of the cases for MSE transrating.

Lagrangian optimization of quantization step-sizes for transrating is presented in chapter 4. This method provides the highest PSNR under given bit-rate constraints as compared with other previously developed requantization algorithms. In this chapter we also report the results obtained when we added to it the MSE and MAP estimation methods introduced in chapter 3.

Chapter 5 proposes a novel extension of the Lagrangian optimization method by applying modifications to the quantized DCT indices. The proposed method outperforms all currently known transrating approaches. To reduce the algorithm complexity, we provide a low complexity trellis-based optimization scheme, and discuss other complexity reduction means as well.

Chapter 6 overviews different methods for image segmentation. An extension of the Lagrangian-based methods of chapters 4, 5, which takes HVS properties into account during transrating, is proposed and tested.

Chapter 7 concludes the thesis, summarizes results and discusses future research directions for transrating.

Chapter 2

Reference Transrating scheme

When building a transrating system, it is necessary to decide on its architecture, and to develop a rate-control scheme. The merit of the final system is judged on the basis of how good it meets the rate constraints, and how good is the video quality of the transrated stream. However, the system complexity is also an important issue when choosing a particular solution. Section 2.1 presents existing transrating architectures. In section 2.2 different rate control schemes are described. In section 2.3 we propose a simple complexity-based transrating scheme that will be used as a reference, to which more advanced methods will be compared.

2.1 Transrating architectures

Following [3],the first, and the most straightforward transrating approach is Cascaded Pixel Domain (CPD) transrating. It consist of a standard decoder and encoder, which are applied one after another, as shown in Fig.2.1. The video is fully decoded and then encoded to a new rate as if working on a source sequence. This scheme has the

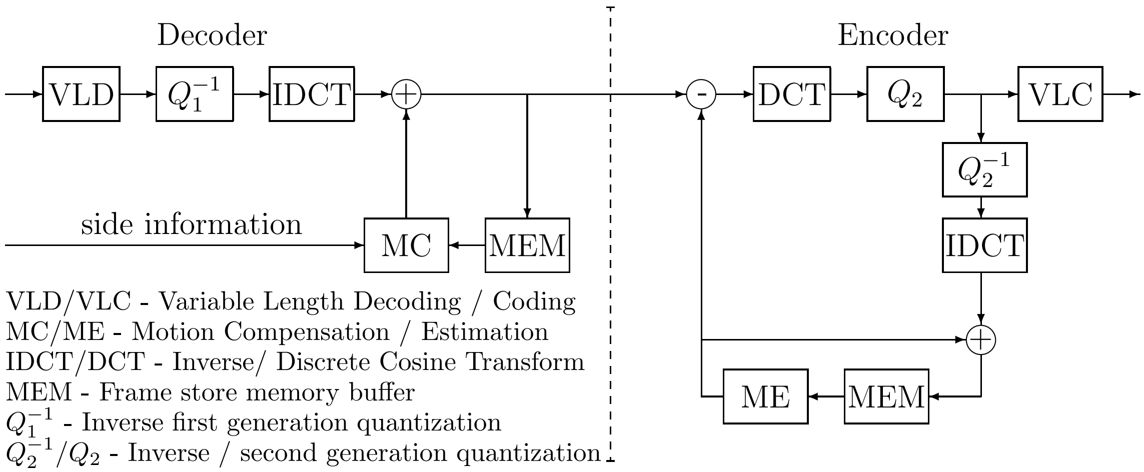


Figure 2.1: Cascaded Pixel-domain Transrating scheme.

following disadvantages:

1. Using the decoded video as a source sequence that is re-encoded compounds the encoding errors and the quality degrades more than with most other compressed-domain schemes.

2. Motion estimation is the most computationally expensive part of the encoding procedure. Since motion estimation was done already during initial encoding, it could be re-used. However, in this scheme this information is not used at all.

3. Since the encoder includes a decoder in its feedback loop, there is doubling of some functional blocks. It increases therefore the overall implementation cost of this scheme.

The most simple scheme is known as Open-Loop (OL) DCT Domain Transrating and is shown in Fig.2.2. It preserves initial encoding decisions as much as possible, and only changes the encoded DCT coefficients by requantization [8] or coefficients

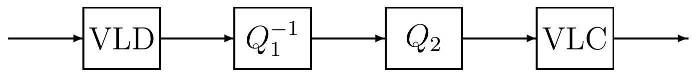


Figure 2.2: Open-loop DCT-domain Transrating scheme.

cropping [16, 17]. This scheme is much simpler and thus very fast. Its main disadvantage is an error drift from reference frames to frames that depend on them.

A Fast Pixel Domain (FPD) [8] Transrating scheme is shown in Fig.2.3. It comes to improve the CPD scheme. It reuses the MB type's decisions as well as initial Motion Vectors (MVs), and thus greatly reduces the encoding complexity. It also utilizes the linearity of motion compensation and of DCT/IDCT to reduce the multiple blocks in the CPD scheme. It is important to mention that while CPD Memory buffers are saving pixel domain predicted frames, FPD Memory buffer (MEM) content is reconstructed requantization error that has to be added to DCT dequantized coefficients to avoid error drift.

FPD transrating needs to return to the pixel domain in order to apply motion compensation on the requantization error. However, Motion compensated DCT blocks can be calculated directly in the DCT Domain as the sum of up to four source DCT blocks multiplied by appropriate DCT-domain shift-matrices [10]. Hence, a DCT Domain (DD) Transrating scheme can be built as shown in Fig.2.4. It was shown in [10] that the number of calculations needed in this case is lower than when applying IDCT, motion compensation and then DCT. However, existing DCT/IDCT accelerators make this solution attractive only for strong hardware development companies.

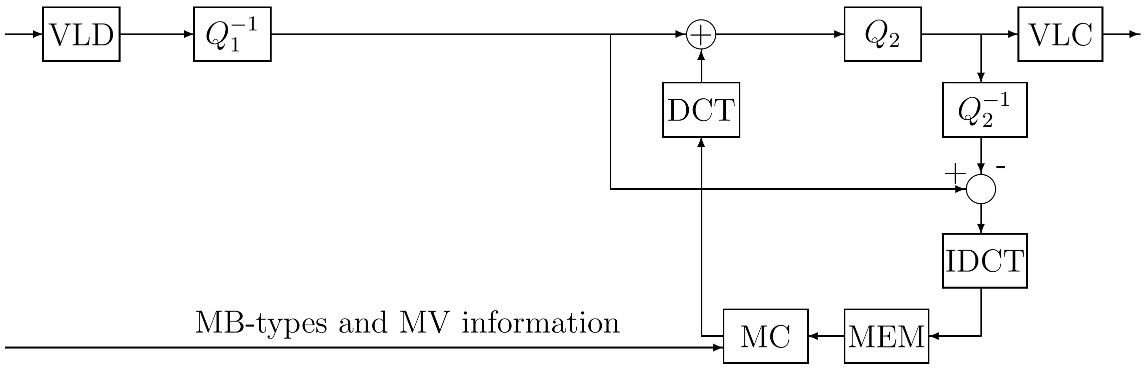


Figure 2.3: Fast Pixel-domain Transrating scheme.

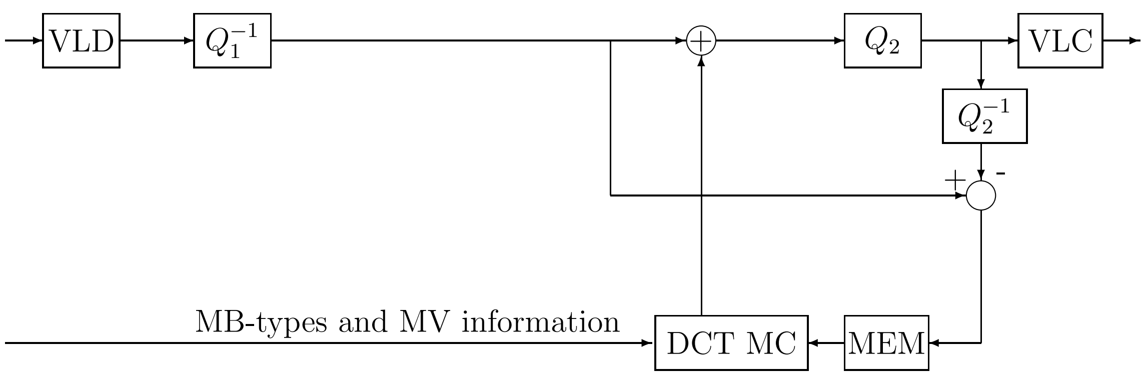


Figure 2.4: DCT-domain Transrating scheme.

2.2 Dynamic Rate-Control for Transrating

There are several methods for dynamic bit-rate control. All of them are based on tracking the available bit-rate status and updating the quantization scale accordingly. Rate control can be separated into two steps: Picture layer bit-allocation, which determines the target bit-budget for each picture, and macroblock layer rate-control that determines the quantization parameters for coding the macroblocks. Sometimes picture level bit-budget allocation can be downscaled to slice level. The simplest frame-level scheme [21] proposes to divide the new bit-budget among frames using the same ratio that they get in the input stream:

$$\frac{R_{out}}{R_{in}} = \frac{T_{out}}{B_{in}} = const \tag{2.1}$$

where:

R_{out} - desired output average bit-rate

R_{in} - input sequence average bit-rate

B_{in} - bits spent on current frame in input stream

T_{out} - target bit allocation for the transrated frame

This approach is straight-forward. It does not provide any problem with virtual buffer fullness - the buffer size can be simply decreased by the same ratio the stream rate goes down. But the output stream quality is not the best possible one because quality reduction of the I-pictures and the P-pictures impacts total video quality more than it does for the B-frames. So another approach [21] proposes to provide the same reduction ratio for all frames as in Eq.(2.1), but to use for I-frames the square root of the ratio of output/input bit-rates; for P-frames - the above ratio itself, and for

B-frames the ratio that will adjust the GOP bit-budget to the desired rate:

$$T_B = \frac{\frac{R_{out}}{R_f}(1 + N_P + N_B) - S_{out,I}(i-1) - S_{out,B}(i-1) - \frac{N_P}{n_P} S_{out,P}(i-1)}{\frac{R_{in}}{R_f}(1 + N_P + N_B) - S_{in,I}(i-1) - S_{in,B}(i-1) - \frac{N_P}{n_P} S_{in,P}(i-1)} \tag{2.2}$$

where:

T_B	- bit-budget for current B-frame
N_P, N_B	- number of P- ,B- frames in GOP
n_P	- number of already transrated P- frames in GOP
R_{out}, R_{in}	- designated output or input bit rates
R_f	- frame-rate
$S_{out/in,I/P/B}(i-1)$	- bit-budget allocated by already transrated I-,P-,B-frames in input/output bitstreams

Picture layer bit-budget allocation in a GOP based on picture complexity is proposed in [22]. Picture complexity is defined as the product of the average quantizer step-size and the number of bits generated, divided by some empirical constant. To estimate the current picture complexity, the ratio of output/input complexities in the previously transrated frame is multiplied by input complexity of the currently transrated frame:

$$X_I = S_I Q_I, X_P = \frac{S_P Q_P}{K_P}, X_B = \frac{S_B Q_B}{K_B} \tag{2.3}$$

where

S_I, S_P, S_B	- number of bits generated by encoding of I-,P- and B-frames, respectively
Q_I, Q_P, Q_B	- average quantization step sizes used in encoding
K_P, K_B	- universal constants (1.0 and 1.4)

The allocation of the current frame-budget in the GOP is proportional to its complexity, for example, for I-frames:

$$T_I = \frac{\frac{X'_{out,I}}{X'_{in,I}} X_{in,I}}{\frac{X'_{out,I}}{X'_{in,I}} X_{in,I} + \sum_{i=1}^{N_p} \frac{X'_{out,P}}{X'_{in,P}} X_{in,P}(i) + \sum_{i=1}^{N_B} \frac{X'_{out,B}}{X'_{in,B}} X_{in,B}(i)} T_{GOP} \tag{2.4}$$

where

T_I	- bit-budget for I-frame.
T_{GOP}	- bit-budget for the whole GOP
$X'_{out,[I/P/B]}$	- average output complexity of transrated I/P/B frames in previous GOP
$X'_{in,[I/P/B]}$	- average input complexity of transrated I/P/B frames in previous GOP
$X_{in,[I/P/B]}$	- input complexity of frames in current GOP

Bit-budgets for P- and B-frames are calculated in a similar way. For open-loop schemes, it was proposed to add a weighting factor to the ratio of averaged input/output complexities which is the square root of relative picture depth [23]. There is no clear explanation about how that depth is defined. Another simplification to Eq.(2.4) is proposed in [24],[25]:

$$T_n = \frac{\frac{X'_{out,n-type}}{X'_{in,n-type}} X_{in,n}}{X'_{out,I} + N_P \frac{X'_{out,P}}{K_P} + N_B \frac{X'_{out,B}}{K_B}} T_{GOP} \tag{2.5}$$

where

T_n	- bit-budget for n-th frame
T_{GOP}	- bit-budget for the whole GOP
$X'_{out,[I/P/B]}$	- average output complexity of transrated I/P/B frames that are transrated already
$X'_{in,[I/P/B]}$	- average input complexity of transrated I/P/B frames that are transrated already
$X_{in,n}$	- input complexity of the current frame
$N_{[P/B]}$	- Number of P/B frames in current GOP

In case the transrating mechanism does not provide the exact number of bits for the particular frame, the same work [25] proposes to modify the bit-rate control to allocate the bits left in the current GOP between the frames left according to there complexities similar to Eq.(2.4). The main drawback of complexity-based methods is the lack of virtual buffer status verification. To take care of it, a buffer-controlled scheme is proposed in [26]:

$$T_n = \begin{cases} \max\{R_{out}, 0.9B_s - B_n\} & \text{if} \quad B_n + T_{1,n} > 0.9B_s \\ Bits_r - B_n + 0.1B_s & \text{if} \quad B_n + T_{1,n} - Bits_r < 0.1B_s \\ T_{1,n} & \text{otherwise} \end{cases} \tag{2.6}$$

$$T_{1,n} = \frac{B_n+2(B_s-B_n)}{2B_n+(B_s-B_n)} \max\{R_{\mathrm{out}}, 0.95\frac{Bits_l}{N_l} + 0.05B_{n-1}\}$$

$$B_n = B_{n-1} + Bits_{n-1} - Bits_r$$

where

T_n - final bit allocation for current frame corrected
to avoid buffer underflow or overflow

$T_{1,n}$ - initial bit allocation according to GOP bit - budget left
vs. number of frames left to transrate

B_s - buffer size

B_n - buffer fullness

$Bits_n$ - bits spend to transrate n - th frame

$Bits_r$ - bits removed from the buffer during current frame
supposed decoding

$Bits_l$ - bits left for current GOP encoding

N_l - frames left to encode in current GOP

R_{out} - Current frame proposed output bit allocation
(bps divided by number of frames per second)

This scheme of bit-allocation according to current buffer fullness helps to avoid overflow or underflow of the buffer. But basically it assumes that the bit allocation for every frame is the same, which is not true for MPEG where I-frames need more bits than P-and B-frames.

2.3 Simple complexity-based transrating

The goal of the current research is to develop a high-quality transrating algorithm. Hence, we have decided to implement a closed-loop transrater. Theoretically, the best choice is to use a DCT-domain transrater. However, although a Fast Pixel-Domain scheme needs slightly more calculations, it is simpler to implement. As we see no place for our contribution in this part of the system, we decided to use the FPD transrating architecture. GOP-level rate-control takes a lot of attention in recent works, but providing the best quality with matching rate and virtual buffer constraints at the same time is still an open issue. We choose therefore a simple solution that doesn't provide problems with virtual buffer fullness: the bit-budget reduction is according to Eq. (2.1). The last thing we have to decide on is how to implement the frame-level bit-rate control. This is what we are going to present in this section. It is possible to see that Eq.(2.4),(2.5) provide a bit budget to every frame of the same type according to its input complexity $X_{in,n}$. To further simplify the problem, we assume that complexity reduction is the same for all kinds of MBs in the same frame. In this case the ratio of input complexities must remain the same, and we can estimate the output bit-rate of MBs left to encode by bit-allocation of the last encoded MB, and update the quantization step-size accordingly:

$$q_{2,n+1}^{ind} = \begin{cases} q_{2,n}^{ind} + 1 & if \quad \hat{B}_{out} > B_{out} \\ q_{2,n}^{ind} - 1 & if \quad \hat{B}_{out} < B_{out} \\ q_{2,n}^{ind} & if \quad \hat{B}_{out} = B_{out} \end{cases} , \ \hat{B}_{out} = \frac{\sum\limits_{n+1}^{N} X_k}{q\_tab[q_{2,n}^{ind}] \cdot B_n} \tag{2.7}$$

where

B_n	-	bit allocation of the last transrated n-th MB
$\hat{B}_{out}$	-	estimated bit number needed to encode MBs left
B_{out}	-	bits left to encode the rest of the frame
$q_{2,n}^{ind}$	-	index of last quantization step-size used
$q\_tab[\,]$	-	quantization step-size table used to get appropriate quantization step-size value

2.4 Experimental results

In section 2.3 we proposed a very simple complexity-based transrating scheme that will be used as a reference for comparing the performance of other, more complex, transrating schemes that will be considered in this work. We use 45 frames of 8 standard video sequences for our analysis and comparisons:

Sequence	format	comments
container	CIF(352x288)	very low motion
foreman	CIF(352x288)	moving face
hall monitor	CIF(352x288)	very stable with small motion
tennis	SIF(352x240)	zoom, motion
coastguard	CIF(352x288)	medium motion
garden	SIF(352x240)	panning, scene depth
football	SIF(352x240)	fast motion, panning
mobile	SIF(352x240)	synthetic scene, medium motion, panning

Table 2.1: Description of video streams used for comparison

Each sequence was encoded into a 4Mbps MPEG-2 stream, using a standard TM5 encoder, and further transrated by each of the methods considered in current work. The sequence rate was reduced to 3Mbps, 2.5Mbps, 2Mbps, 1.5Mbps, 1Mbps, which provides rate reduction factors from 1.33 to 4. Additionally, each stream was re-encoded by full decoding and encoding using a standard encoder to the reduced rates, and each original sequence was also directly encoded to these rates.

Figures 2.5 and 2.6 present PSNR results as a function of rate for the reference simple transrating scheme (Sim) , original sequence encoding (Enc), and cascading of decoder and encoder (Re) for each one of the above sequences. It is possible to divide the sequences into two classes:

a) Container, HallMonitor, Foreman.

b) Mobile, Football, Tennis, Garden, CoastGuard.

In the first class all graphs have higher PSNR values than those of the second class. Simple transrating outperforms cascading up to nearly 4 times rate reduction, and can even outperform direct encoding for rate-reduction factors close to 2. Encoding and cascading graphs do not preserve convexity, possibly because those sequences do not match TM5 rate-control assumptions.

In the second class, although all sequences have lower PSNR values, they remain convex. Simple transrating is always worse than encoding of the original sequence, but outperforms re-encoding up to rate reduction factors close to 2.7 (1.5 Mbps).

Both classes show the importance of utilizing of original encoding decisions as long as the rate reduction factor is not too high. For high rate reduction factors, old motion information is not good any more because the change in reference frames is too big, and re-estimation of motion provides better results.

As for other transrating approaches, we observed the same two classes as mentioned above, so we will show in the sequel only the results for the FOREMAN sequence - from the first class, and for GARDEN and TENNIS sequences - from the second class.

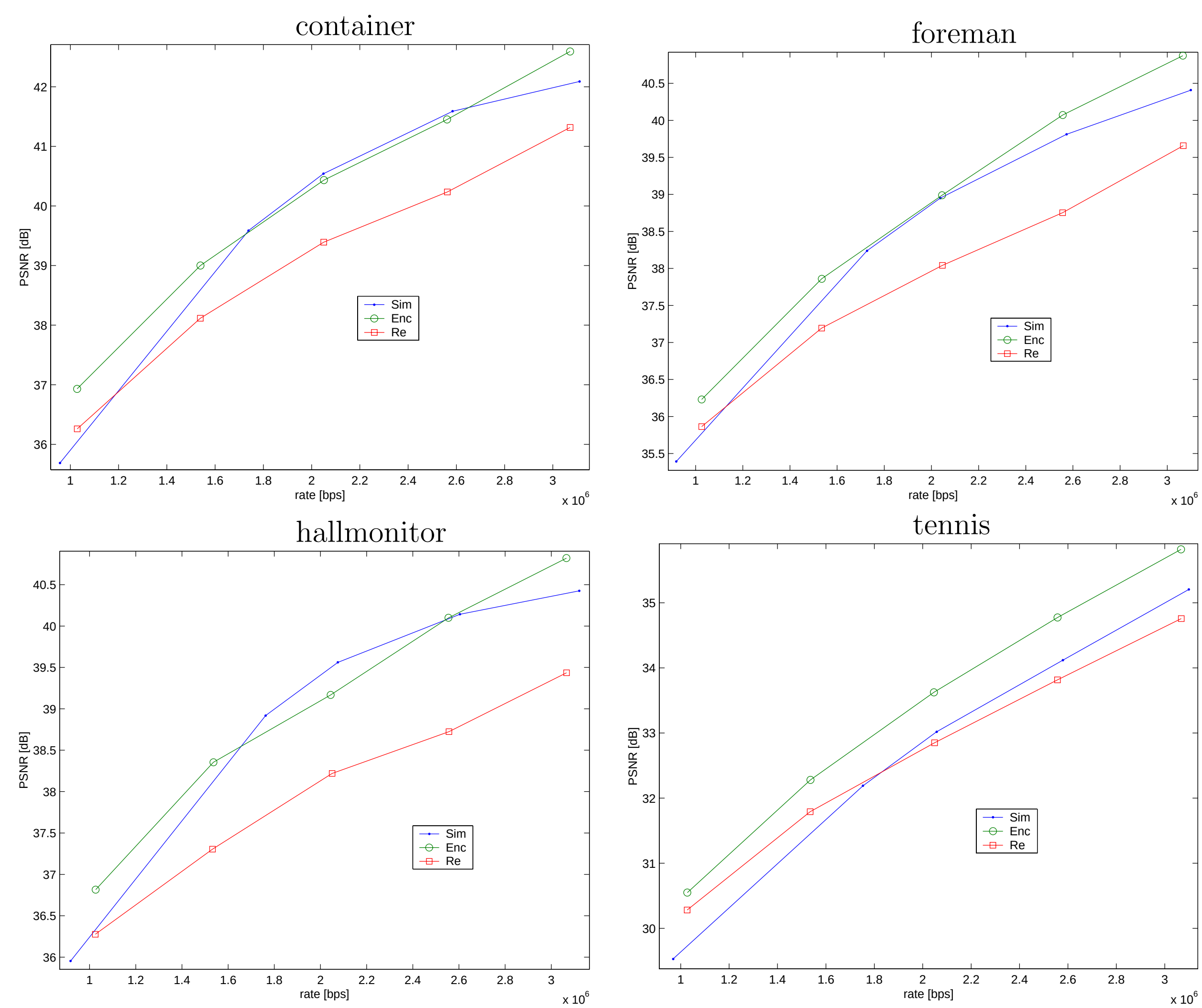


Figure 2.5: Simple transrating compared with TM5 encoded and Re-encoded.

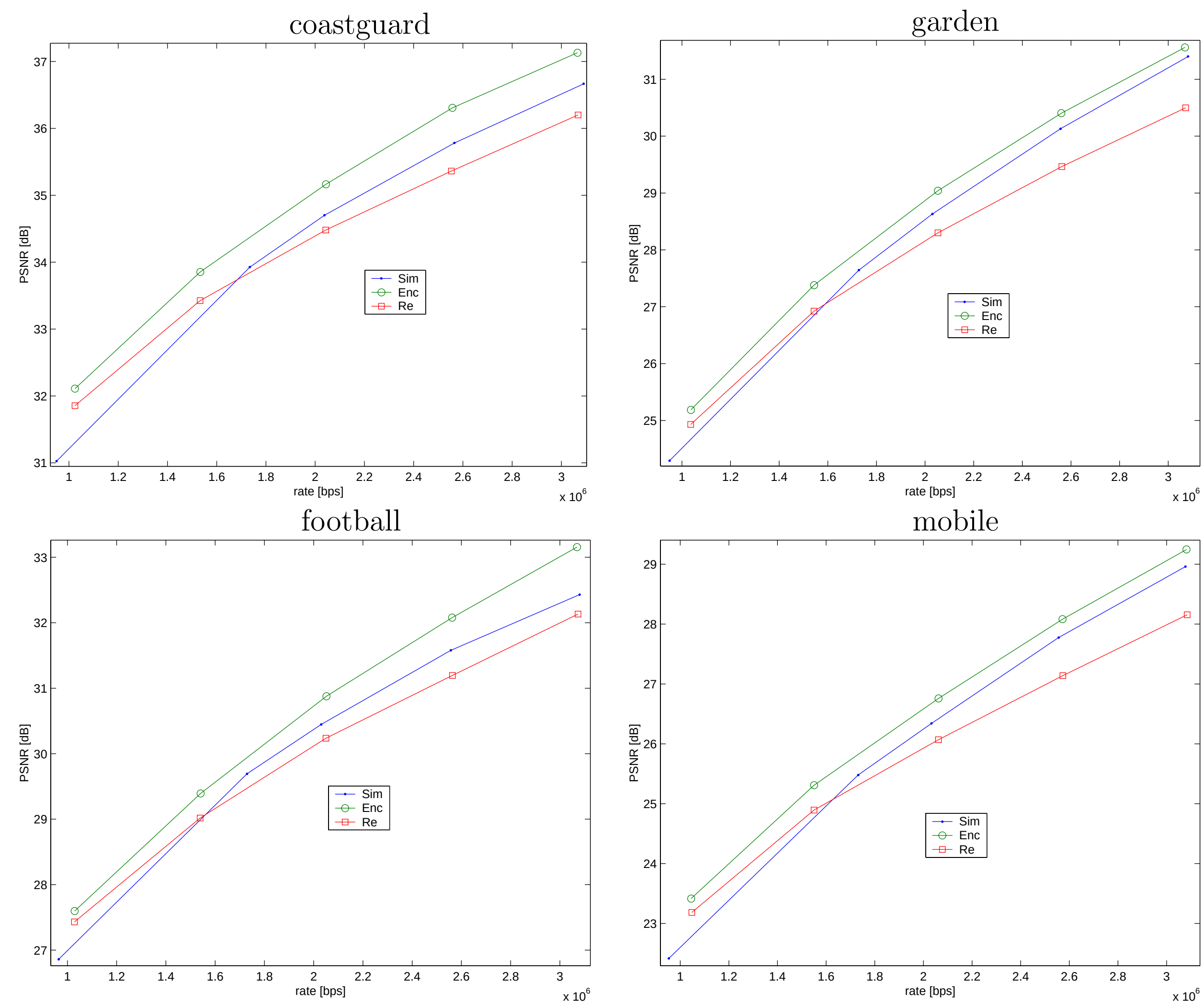


Figure 2.6: Simple transrating compared with TM5 encoded and Re-encoded.

Chapter 3

Requantization based on DCT coefficients distribution

3.1 Cascading-error problem

MPEG-2 defines the set of representation levels for a given quantization step and a DCT coefficients weighting matrix, which coarsely accounts for non-uniform sensitivity of the HVS to different spatial frequencies. Still there is a possibility for every encoder or transcoder to choose the particular decision interval for every representation level. It is obvious that for efficient encoding the decision intervals should be set to minimize the average or maximum distortion. Thus, the partitioning depends on the probability distribution function of each DCT coefficient (except DC, which is treated separately). In a transrating application, we get the set of values of DCT coefficients distributed discretely over representation levels of particular quantization steps. Because of the error introduced in the initial quantization, requantization may produce a cascading error that is larger than the error obtained by direct encoding

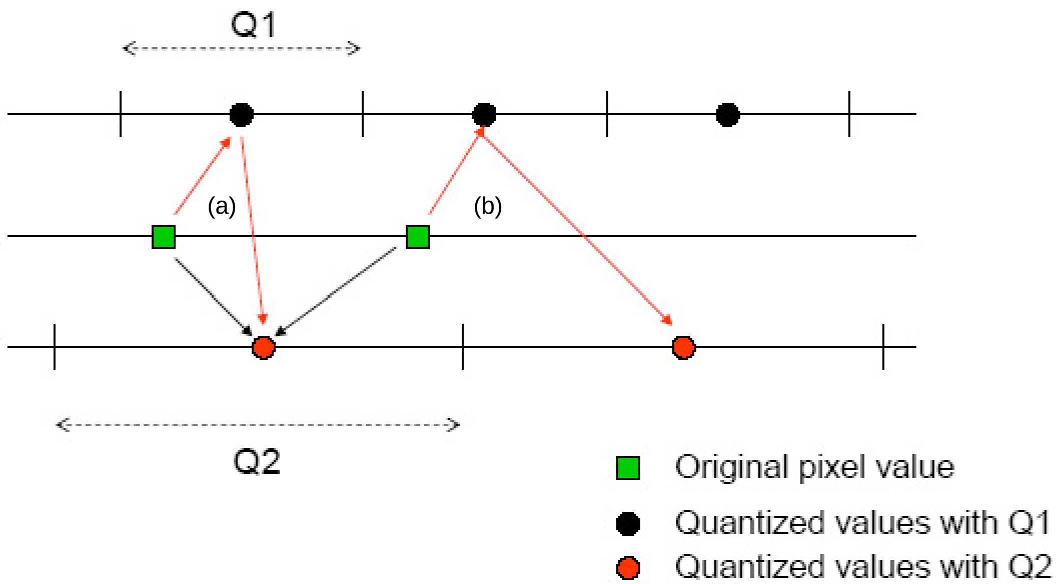


Figure 3.1: (a) no cascading error introduced (b) cascading error due to requantization is greater than error of direct quantization.

with the same quantization step, as described in [4] and shown in Fig. 3.1.

A method for reducing the cascading error is proposed in [29]. It is based on the assumption that all decision regions are uniformly distributed and symmetric. Two terms are proposed. The first is termed the critical requantization ratio, between initial and final quantization steps, that may result in the biggest possible cascading error, like $Q_2/Q_1 = 2$ in the example below. The second is termed the optimal ratio as it is a ratio that doesn't cause cascading errors at all, like $Q_2/Q_1 = 3$ in the following example. The idea is, of course, to avoid critical ratios and, furthermore, if the ratio proposed by the bit-rate control mechanism is near the optimal one, the proposed quantization step size is changed to yield an optimal ratio. Fig. 3.2 gives examples of optimal and critical requantization ratios, assuming a uniform midstep quantizer with no dead zone (also known as midtread quantizer - see Fig. 3.5(a)). For the critical ratio, $Q_2 = 2Q_1$, where Q_1, Q_2 denote the original and transrating

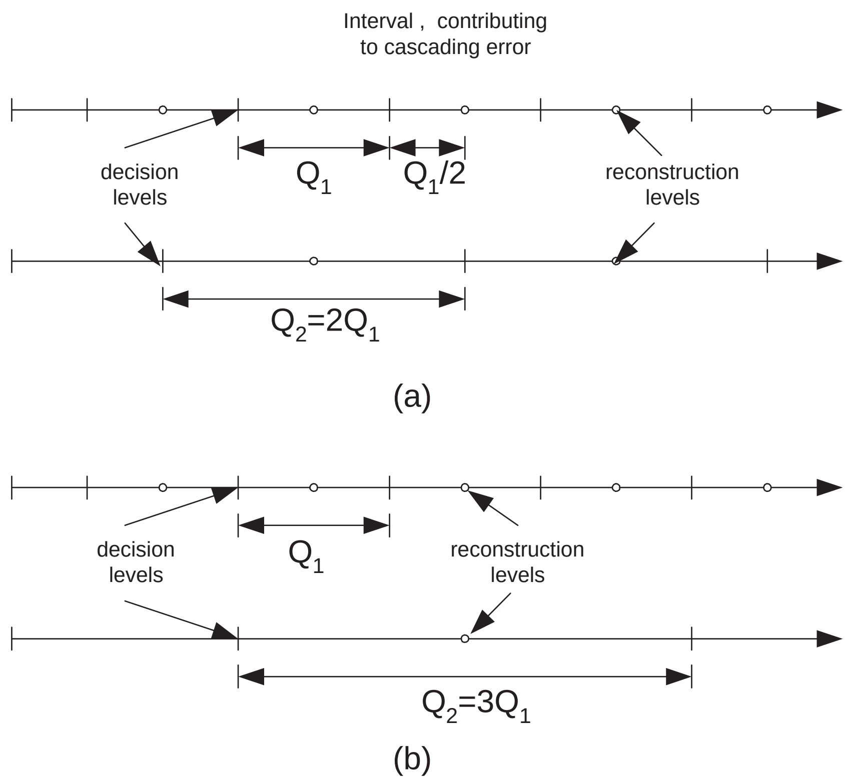


Figure 3.2: (a) range supporting cascading error is maximal (b) no cascading error introduced.

quantization step-sizes, respectively, every decision level defined by Q_2 will always be equal to one of the representation levels defined by Q_1. Thus, the range of values that will be requantized to a different value than that of direct quantization will be equal to $Q_1/2$ for every representation level produced by Q_2, as shown in Fig. 3.2(a). For an optimal ratio, $Q_2 = 3Q_1$, every decision level defined by Q_2 will always be equal to one of decision levels defined by Q1 and no cascading error will be introduced, as seen in Fig. 3.2(b). One of the first approaches that take the distribution of transform coefficients into account is proposed in [6]. A parametric rate-quantization model, based on traditional rate-distortion theory, is applied to MPEG encoders. Given the bit-rate budget for a picture, this model calculates a baseline quantization scale factor. For a rate-distortion analysis, the coefficients are assumed in [6] to be Gaussian random variables, and the distortion to be proportional to the square of the quantization step.

The variance of every transform coefficient in intra-frames was estimated and so the distortion and rate for every quantization step is calculated by:

$$R = r_1 + \sum_{i=2}^{N} \frac{1}{2} \log_2 \frac{\sigma_i^2}{d_i}, \;\; D = d_1 + \sum_{i=2}^{N} d_i \tag{3.1}$$

where,

N	- number of transform coefficients
σ_i^2	- estimated variance of i-th coefficient
d_i	- quantization distortion (proportional to Q_i^2)
d_1, r_1	- distortion and bit-rate of DC coefficient
R	- bit-rate estimation
D	- estimated distortion

MPEG-2 provides different encoding of DC coefficients in intra-blocks than for the rest of DCT coefficients, so the bit-rate and distortion for DC have to be calculated separately. Assuming that the distortion is proportional to the square of the quantization step-size and of the appropriate factor from quantization matrix, [6] uses eq.(3.1) to estimate the output bitrate. One step-size for the whole frame, which provides the closest rate to the bitrate constraint, is chosen. In a similar way the procedure is applied for non-intra-frames and is used for optimal quantization step estimation at the picture-level.

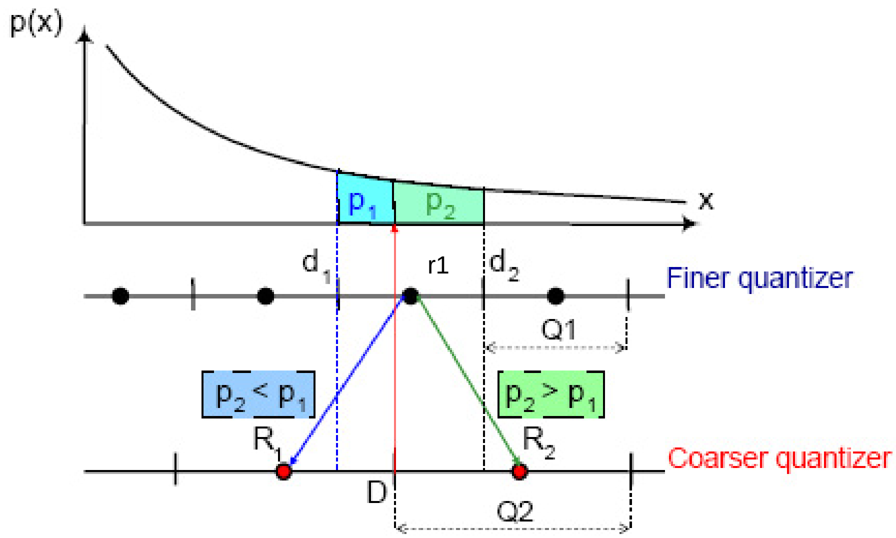


Figure 3.3: MAP requantization decision is taken based on probabities p_1 and p_2.

3.2 MAP and MSE requantization

To address the problem of cascading error described in section 3.1, it is necessary to take into account the probability distribution of input values. The idea of Maximum A-Posteriori (MAP) requantization was introduced by Werner in [30]. Instead of checking quantization ratios without questioning the optimality of the decision intervals, as in section 3.1, the MAP approach minimizes the average cascading error by an optimal mapping from a given set of initial encoding representation levels to the defined requantization representation levels. The problem can be seen as changing the decision intervals on requantization scale for each pair of initial and requantization step sizes. The quantization step-size selection defines particular representation levels for initial quantization and requantization. The MAP requantization method aims at minimizing the probability of representing original input values by a representation value different from the value that would be assigned to it by direct quantization. Fig. 3.3 presents an initial pdf function $p(x)$ and initial quantization and requantization decision and representation levels. r1 is the representation level of decision interval $[d_1, d_2)$. The initial quantization scale is defined by quantization step-size Q1 and the lower boundary of the decision interval d_1. The requantization scale (without MAP) is defined by Q2 and the corresponding boundary D. This scale is denoted as 'reference'. The decision to be taken by MAP quantization is whether the values in $[d_1, d_2)$ should be mapped to R_1 or to R_2. For this purpose the probability that the values that are mapped to R_1 come from the interval $[d_1, D)$ is compared with the probability that they come from the interval $[D, d_2)$. If the first probability is bigger, than interval will be mapped to R_1, otherwise to R_2.

The MAP decision rule minimizes the cost function that represents how much MAP quantization is different from the direct encoding decisions:

$$\mathrm{E}\left[\left(y_{2,ref} - y_2 \right)^2 | y_{1,ref} \right] \tag{3.2}$$

Assuming that the transform coefficients pdf is Laplacian with parameter α:

$$p(x) = \frac{\alpha}{2} e^{-\alpha|x|}, \tag{3.3}$$

where α is related to the standard deviation σ via $\sigma = \frac{\sqrt{2}}{\alpha}$. It is claimed in [16] that this is an appropriate pdf for DCT coefficients of Intra frames. The MAP rule in this case becomes:

$$\begin{aligned} &y_2 = Q_2(y_1 = r1) = \begin{cases} R_1 & , \frac{D-d_1}{q_1} > \frac{v_1}{2} \\ R_2 & , \frac{D-d_1}{q_1} < \frac{v_1}{2} \end{cases} \\ &v_1 = \frac{1}{\alpha Q_1} \ln\left(\frac{2}{1+e^{-(\alpha Q_1)}} \right) \end{aligned} \tag{3.4}$$

This mapping can be calculated offline with some assumptions on the initial pdf, or estimated online from particular video stream statistics, as will be described in the

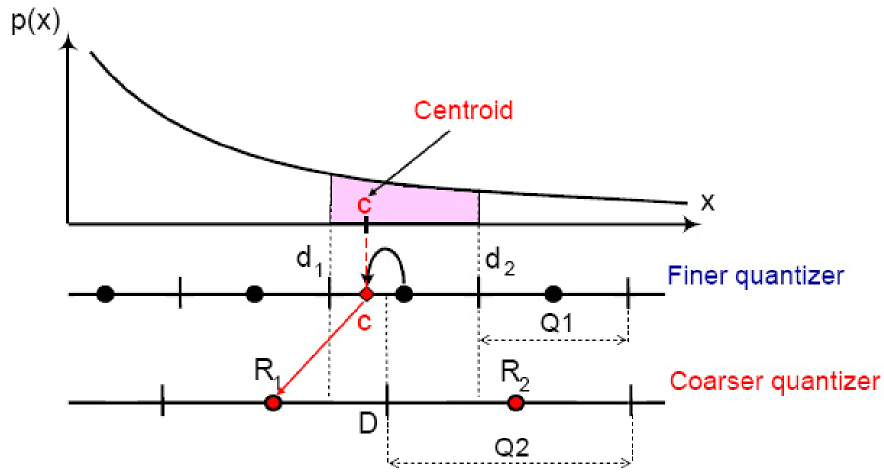


Figure 3.4: MSE requantization decision is taken based on centroid position.

following section.

Another approach is known as MSE minimization. Assuming $p(x)$ and boundaries of the initial decision region are known, the centroid of that area is the best choice for the representation level that can be further quantized by a new quantization scale as shown in Fig. 3.4. The MSE decision rule minimizes the cost function that represents how much MSE quantization is different from the direct encoding decisions:

$$\mathrm{E}\left[(x-y_2)^2|y_{1,ref}\right] \tag{3.5}$$

The main disadvantage of the methods described above is that they are developed for Intra frames transrating or open-loop transrating, since they do not consider error compensation that should be done if error propagation in a GOP is to be avoided. The above methods are all assuming that values for requantization lie on the initial quantization scale, which is not true in the case of closed-loop error compensation.

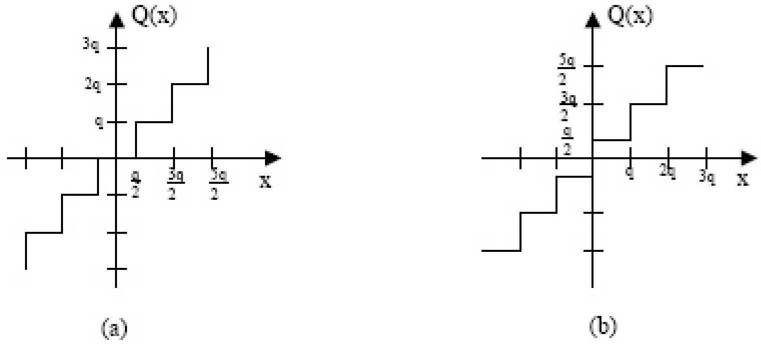


Figure 3.5: (a) midstep quantizer with no dead zone (b) midrise quantizer.

One of the possible extensions of the above results is to analyze the distribution of coefficients in a closed-loop transcoder before requantization and to use it for the optimal definition of decision intervals /representation values.

3.3 Estimation of pdf parameters

MAP and MSE requantization, as any method based on the statistical distribution of the initial DCT values, requires knowledge of the pdf, usually determined by its parameters. In [31] a method for estimating the Laplacian parameter a of the DCT coefficients (see eq. (3.3)) from the encoded coefficients distribution is proposed. It can be estimated for each pdf of every DCT coefficient, in every block of each MB, from the statistics of the current frame, or for each DCT coefficient in the entire frame, while tracking its statistics over time. It is shown in [31] that for a uniform midstep quantizer with no dead zone (Fig. 3.5(a)), the Laplacian parameter α could be estimated by solving the following quadratic equation for z:

$$(A + B + C)z^2 + Bz - A = 0 \tag{3.6}$$

where,

$A = \sum\limits_{l=1, y_l \neq 0}^{N} (y_l - \frac{q}{2})$,with N - number of coeff.,

$B = \frac{N_0 q}{2}$,with N_0 - number of zero coeff.,

$C = (N - N_0)q$,

Then, for from the relation $z = e^{-(\hat{\alpha}\frac{q}{2})}$, the estimated parameter is

$$\hat{\alpha} = -\frac{2}{q}\ln z \tag{3.7}$$

For a uniform midrise quantizer, shown in Fig. 3.5(b), the equation becomes:

$$(A + B)z - A = 0 \tag{3.8}$$

where,

$$A = \sum_{l=1}^{N} (y_l - \tfrac{q}{2}),$$
$$B = \tfrac{Nq}{2}.$$

And the estimation for α is again $\hat{\alpha} = -\frac{2}{q}\ln z$. These results were obtained in [31] by using Maximum-Likelihood (ML) estimation.

Another work [32] proposes a method for estimating the Laplacian parameter α from the variance of the quantized coefficients:

$$v_q = \sum_{n=-\infty}^{\infty} (nq)^2 p_n = q^2 \frac{e^{\alpha q/2} + e^{-\alpha q/2}}{(e^{\alpha q/2} - e^{-\alpha q/2})^2}, \tag{3.9}$$

where,

$$p_n = \int\limits_{nq-\frac{q}{2}}^{nq+\frac{q}{2}} p(x)dx \tag{3.10}$$

is the probability of the original coefficient value to fall into n-th decision interval.

The Laplacian parameter is estimated as follows: From eq.(3.9):

$$e^{\alpha q/2} + e^{-\alpha q/2} = \frac{q^2 + \sqrt{q^4 + 16v_q^2}}{2v_q} = u \tag{3.11}$$

hence,

$$e^{\alpha q/2} = \frac{u + \sqrt{u^2 - 4}}{2} = t, \tag{3.12}$$

resulting in,

$$\hat{\alpha} = \frac{2}{q} \log t \tag{3.13}$$

It predicts the quantization error, for the ideal case of infinite number of representation levels, to be:

$$e_q^2 = \alpha \int_0^{q/2} x^2 e^{-\alpha x} dx + \alpha \sum_{n=1}^{\infty} \int_{nq-q/2}^{nq+q/2} (x - nq)^2 e^{-\alpha x} dx = \frac{2}{\alpha^2}\left(1 - \frac{\alpha q}{e^{\alpha q/2} - e^{-\alpha q/2}}\right) \tag{3.14}$$

The estimation of α from (3.9) is not optimal in any sense, and the assumption of an infinite number of possible quantization levels is not true for real implementations. Still it provides a simple way for estimating α that typically supplies a good result. Yet another method for estimating α is proposed by [33]. It is based on the empirical number of zero valued coefficients and +q and -q valued coefficients. It is worse than the previous method because it does not use all the statistics available for estimation, but at this price it is less expensive computationally.

3.4 PDF Estimation based on Linear Interpolation

All the methods described above have the same common disadvantage: they do not support the case when different quantization step-sizes are applied for different MBs. It was also questioned by a number of recent works if DCT coefficients fit the Laplacian distribution model. As we will see in section 3.5, the estimation results are not very good for real coded sequences. To solve this problem, we developed another method to estimate the PDF distribution from quantized coefficients.

The proposed solution is coming from the class of kernel-based methods, which assume that the pdf $p(x)$ can be represented as:

$$p(x) = \sum_k c_k \cdot \phi(x - s_k, \theta_k) \tag{3.15}$$

$\phi(x)$ is called kernel function, and can be any appropriate positive function such as Gaussian, spline, etc. The preceding model tries to represent the unknown pdf by a linear combination of shifted copies of of the fixed function $\phi(x)$. The shift factor s_k and function parameter θ_k are typically fixed, and the weighting factor c_k is adjusted based on the measurements of the random parameter x.

We used splines as kernel functions. The problem with high order splines is that they don't preserve positivity. The splines we chose to use are zero-order, more known as zero-order hold interpolation, and first-order, which are no more than well-known linear interpolation. As first-order splines gave better results in simulations, only their results will be presented here. In order to simplify the calculations, we decided to work with discrete pdf that is defined for integer points on the positive real axis.

The proposed pdf estimation algorithm for a particular DCT coefficient is very simple:

1. Reset final discrete pdf prediction vector to zero.
2. Get the first quantization step-size used.
3. Build histogram of quantized coefficients (it will be populated only in multiples of quantization step-size).
4. Estimate pdf values at each discrete point by linear interpolation between values of closest quantization step-size multiples.
5. Add resulting pdf estimation to final pdf prediction vector.
6. If quantization step-size checked is the biggest one, exit.
7. Get next used quantization step-size.
8. Goto 3.

In spite of its simplicity, this estimation method outperforms estimations made by Laplacian-based methods as we will see in the following section.

3.5 Experimental results

In this section different MSE and MAP approaches are compared. We implemented Laplacian pdf parameter estimation using ML (exp-ML) and variance (exp-var) methods, as well as Linear interpolation (linear) method proposed in the previous section.

Fig. 3.6 shows that for MSE method Linear interpolation ($\texttt{Sim}_{\texttt{mse-linear}}$) outperforms Laplacian model based methods for rates above 2Mpbs (rate reduction factor up to 2). For higher rate reduction factors, variance estimated Laplacian model ($\texttt{Sim}_{\texttt{mse-exp-ML}}$) gives better performance over other MSE approaches. This is true for

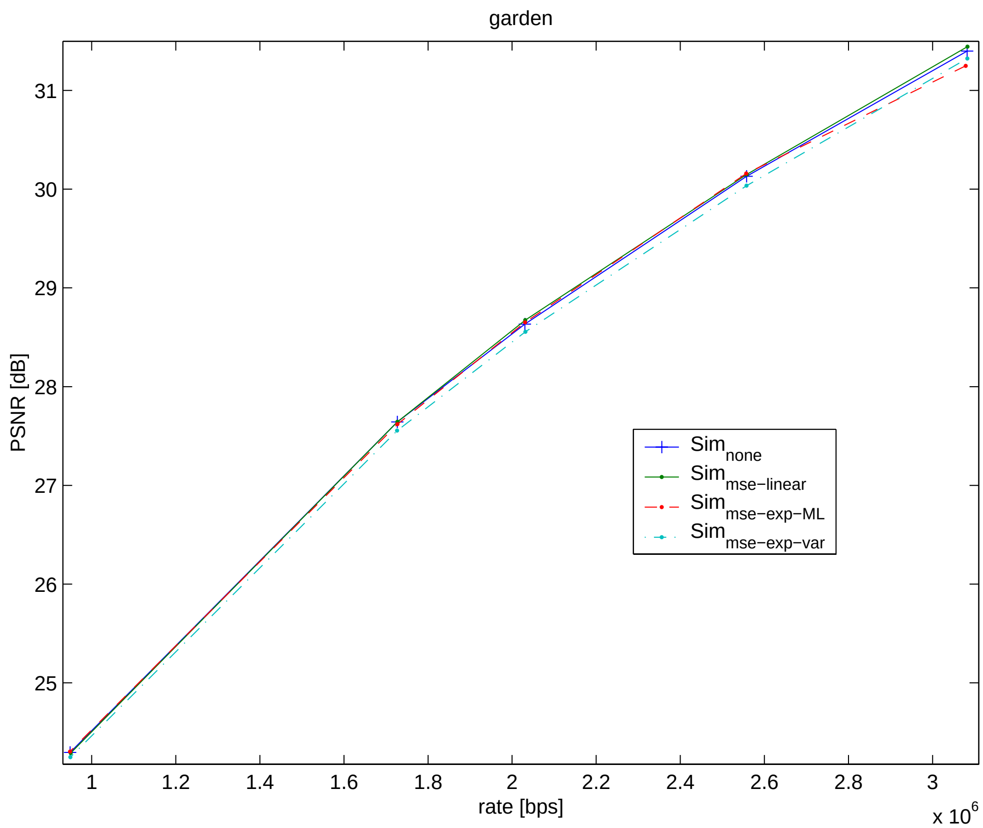
MOBILE, FOOTBALL, TENNIS, GARDEN and COASTGUARD sequences. For other sequences, Linear interpolation is good for high bitrates, but below 2.5 bps ML Laplacian parameter estimation ($\mathtt{Sim}_{\mathtt{mse-exp-var}}$)is the best method.

As for the MAP approach, Laplacian model based methods only reduce the PSNR, while results of linear interpolation ($\mathtt{Sim}_{\mathtt{map-linear}}$) are always above the transrating, which does not use any estimation at all ($\mathtt{Sim}_{\mathtt{none}}$), and become better for higher compression factors, as it is shown on Fig. 3.7.

For GARDEN sequence (and other sequences in the same class) it is better to use MSE for higher rates and MAP for greater compression ratios, with Linear interpolation for pdf estimation in both cases ($\mathtt{Sim}_{\mathtt{mse-linear}}$ and $\mathtt{Sim}_{\mathtt{map-linear}}$), as we can see on Fig. 3.8. The best performance for FOREMAN and others in his class is provided by MSE with linear interpolation pdf estimation ($\mathtt{Sim}_{\mathtt{mse-linear}}$) - for rates above 2Mbps, and MSE with variance Laplacian model pdf estimation ($\mathtt{Sim}_{\mathtt{mse-exp-var}}$) - for higher compression.

MAP method with Linear interpolation can be good for high compression ratios for sequences with relatively high motion. MSE with Linear interpolation seems to work better than Laplacian model if the compression factor is relatively small. The benefit in using the best of MAP or MSE reaches 0.1 dB.

Additional examples can be found in Appendix A.



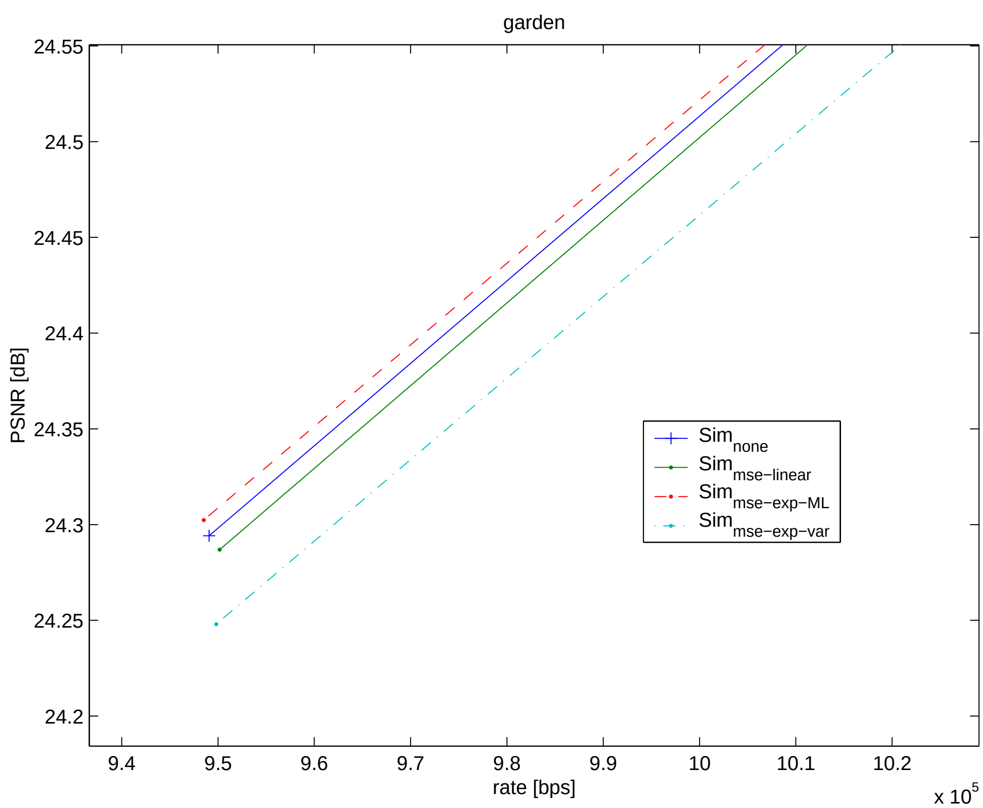


Figure 3.6: MSE simple transcoding, GARDEN sequence.

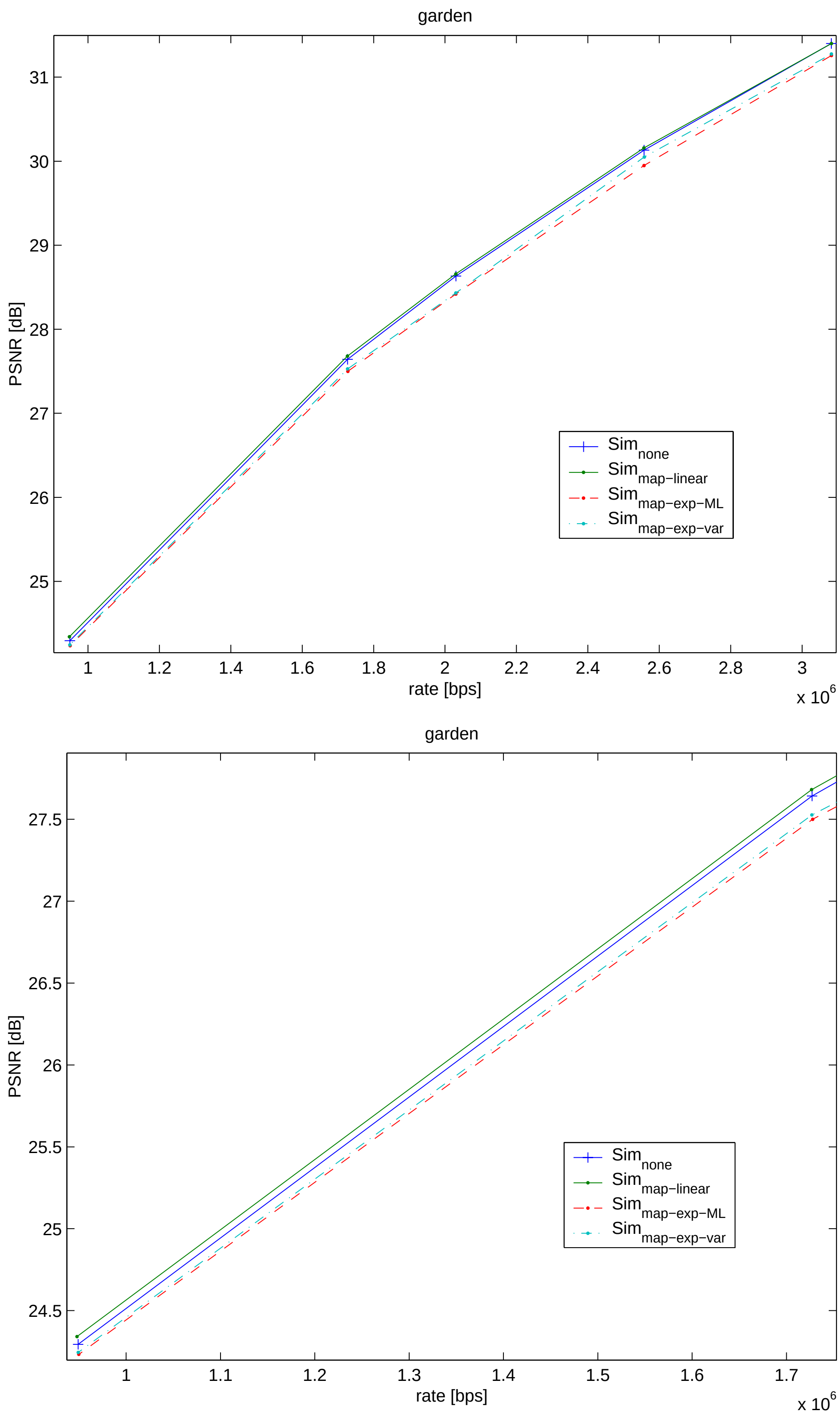
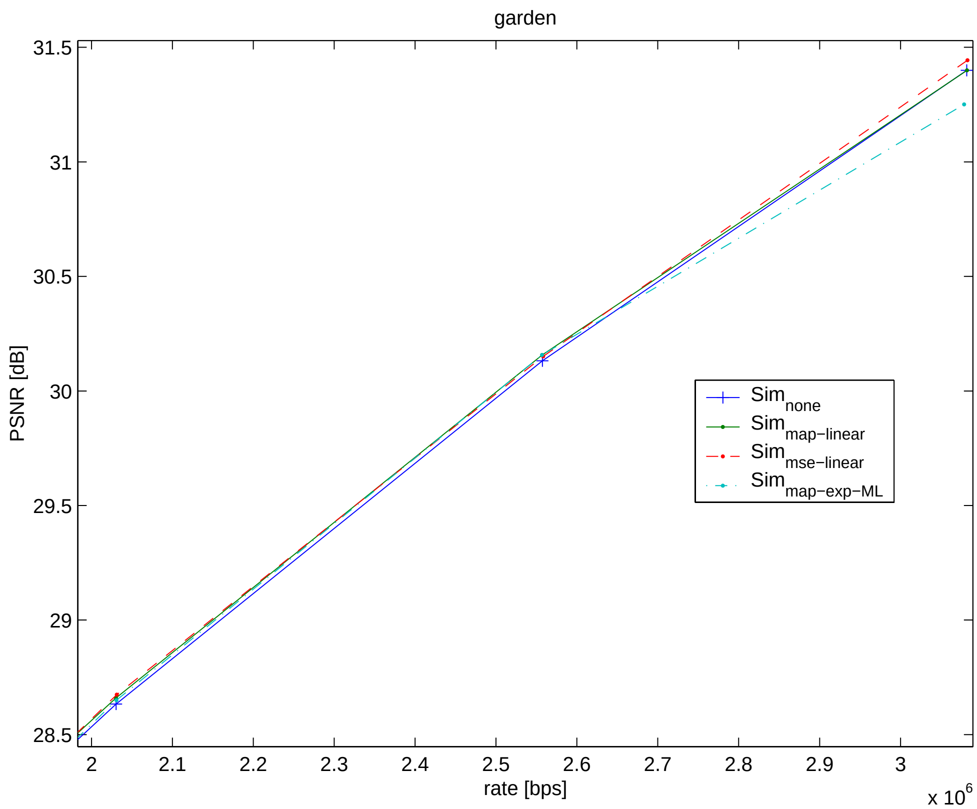


Figure 3.7: MAP simple transcoding, GARDEN sequence.



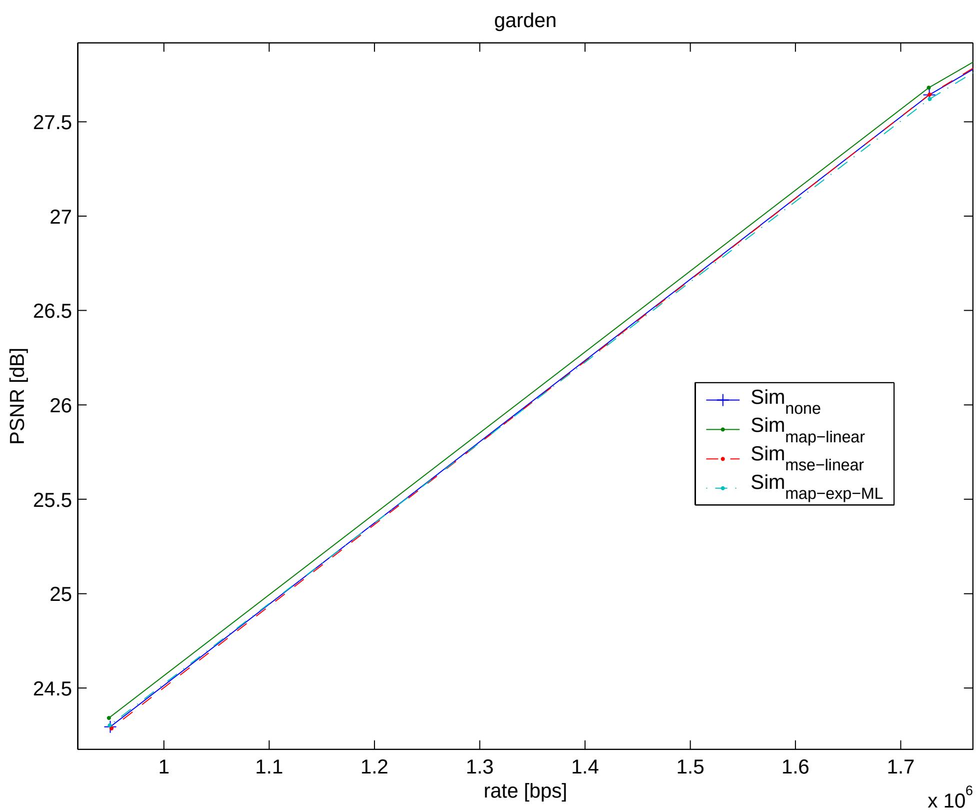


Figure 3.8: compare MSE and MAP for simple transcoding, GARDEN sequence.

Chapter 4

Requantization via Lagrangian optimization

MPEG-2 encoding provides output bit-rate control by changing the quantization step size. The change in quantization step-size enables to achieve bit-rate reduction at the cost of perceptual video quality. There are many questions that arise during transrating, the most important of which are these:

1. How to achieve the desired bit-rate after transrating?

2. How to provide the best perceptual quality for the same rate?

In this chapter we review and examine optimal requantization via Lagrangian optimization. Section 4.1 describes the standard MPEG-2 encoding procedure. Section 4.2 presents the Lagrangian optimization method proposed in [4, 5, 9, 10] to get optimal requantization step-sizes for transrating. Experimental results and their analysis are provided in section 4.3.

4.1 MPEG-2 AC coefficients encoding procedure

Following the application of the DCT to each of 4 luminance 8×8 blocks and to 2 up to 8 chrominance blocks (depending on video format), which form a MB, the DCT coefficients, except for the DC coefficient, are quantized. For each MB, a value from one of two possible tables, each having 32 quantization step-size values, is selected (a different table can be chosen for each frame). The actual quantization step-size used for each coefficient is the product of the selected step-size from the table and a value defined by a suitable quantization matrix that depends on the MB type. The 63 quantized AC coefficients are concatenated in an order defined by one of two possible zig-zag scans. The resulting 6 to 12 vectors, of 63 quantized coefficients each, constituting a MB, are entropy coded by a variable-length-coding (VLC) table. Each coefficient vector is divided into several parts, with each part consisting of a *run* of consecutive zeros followed by a non-zero *level* value, defining a run-level pair. In case of adjacent non-zero level values, the run length is defined to be zero. The MPEG-2 standard defines for every run-level pair a variable-length codeword. There are two VLC tables that can be used. It is possible to use the same table for all types of MBs, or to use a different one for Intra MBs [34].

4.2 Lagrangian optimization

The requantization problem can be formulated as an optimization problem of determining a set of quantization step-sizes that minimize the total distortion in each frame, under a given bit-rate constraint:

$$\min_{\{q_k\}} D, \text{ under the constraint } R \leq R_T \tag{4.1}$$

with ,

$$D = \sum_{k=1}^{N} d_k(q_k),\ R = \sum_{k=1}^{N} r_k(q_k), \tag{4.2}$$

where,

N - number of MBs in the frame;

q_k - quantization step-size for the k-th MB;

d_k - distortion caused to the k-th MB;

r_k - number of bits produced by the k-th requantized MB.

An analysis for the conventional MSE distortion metric is presented in [5]. The problem can be converted into an unconstrained one by merging rate and distortion through a Lagrange multiplier $\lambda \geq 0$ into the cost function:

$$J_{total} = D + \lambda R \tag{4.3}$$

λ defines the relative importance of rate against distortion in the optimization procedure.

This constrains the set of possible solutions to a lower boundary of Convex Hull of all possible solutions. Fig.4.1 shows the constraints of the Lagrangian solution. Points show achievable rate-distortion states. The dotted blue line shows the boundary of the Convex Hull of all possible solutions. Only states on the Convex Hull boundary (blue points on blue dotted line) can be found by Lagrangian optimization. Red states (connected by solid red line), which are optimal for particular rates, can't be reached using Lagrangian optimization. Lagrangian multiplier defines the rotation of the rate-distortion axes before minimization. Under all possible rotations red state will never be the minimum over all possible states. It reduces a bit from the optimality of the solution, but assuming the set of all solutions is dense it should be good enough.

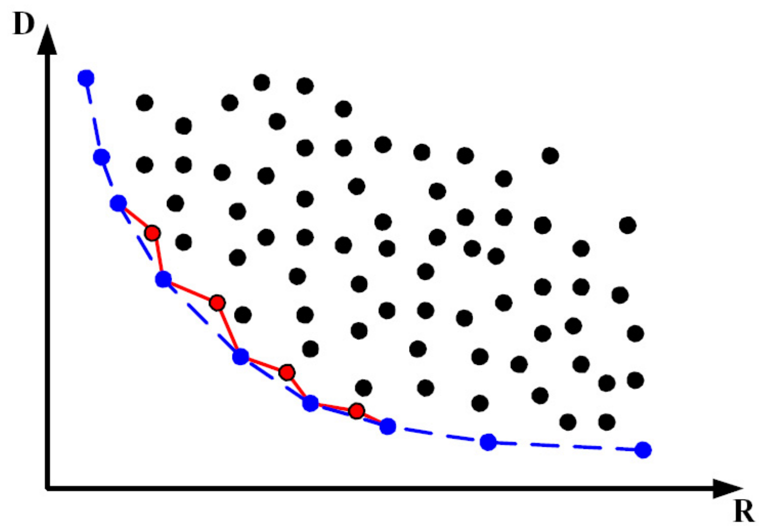


Figure 4.1: Constraints of Lagrangian optimization: only rate-distorion states on Convex Hull boundary (dotted blue line) of all possible states (blue points) can be reached. Optimal states marked by red points, which are connected by solid red line, won't be found.

The Lagrangian cost can be independently calculated for each MB. Thus, for the k-th MB:

$$J_k(\lambda) = \min_{q_k}\{d_k(q_k) + \lambda r_k(q_k)\} \tag{4.4}$$

Let $\{r_k(\lambda), d_k(\lambda)\}_{k=1}^{N}$ be the set of solutions for a particular λ that achieves the minimum Lagrangian cost for every MB. If for a particular value $\lambda = \lambda_s$ the total rate $R(\lambda_s) = \sum_{k=1}^{N} r_k(\lambda_s) = R_T$, than the set $\{q_k(\lambda_s)\}_{k=1}^{N}$ is the optimal set of quantizer step sizes to be used for transrating. λ_s has to be found by some kind of search for every frame if the problem is solved on a picture level, or for every slice, if bit-rate allocation is provided on that level. The proposed transrating distortion measure in

[5] is:

$$d_k(q_k) = \frac{1}{L^2}\sum_{n=0}^{L-1}\sum_{i=0}^{63}\begin{cases} [C_k^{ni}(q_k) - C_k^{ni}(q_k')]^2 & \text{(intra)} \\ [C_k^{ni}(q_k) - C_k^{ni}(q_k') + X_k^{ni}]^2 & \text{(forward,backward)} \\ [C_k^{ni}(q_k) - C_k^{ni}(q_k') + \frac{X_k^{ni}+Y_k^{ni}}{2}]^2 & \text{(interpolated)} \end{cases} \tag{4.5}$$

where,

$C_k^{ni}(q_k)$	- i-th inverse quantized DCT coeff. of block n in MB k
$C_k^{ni}(q_k')$	- same as $C_k^{ni}(q_k)$, but after requantization
X_k^{ni}, Y_k^{ni}	- error drift from reference pictures (forward and/or backward)
L, n	- number of blocks in MB and block index
k	- index of MB
i	- DCT coeff. index (64 DCT coeff in every block)

It is shown in [5], by simulations, that by using the same set of coding decisions as produced by the initial encoder, optimal requantization can achieve higher PSNR than that achieved by direct encoding of the original video sequence to the final lower bit-rate by the usual encoder. This is possible because common encoders do not perform this kind of optimization, for finding the best possible quantization steps, to avoid an increase in the encoding complexity.

Lagrangian optimization is a general and very powerful method. It can also be applied as in [35] to obtain the optimal discarding of DCT coefficients in I-frames. In [17] it is applied to the discarding of high-frequency DCT coefficients only.

4.3 Experimental results

In this section Lagrangian optimization performance and the addition of different statistical methods are presented. As in section 2.4, MSE and MAP approaches based on different pdf estimation methods, which are described in chapter 3, are compared. The additional value of those methods is compared with the improvement provided by the Lagrangian optimization approach.

A simple bi-section search algorithm, which iteratively updates the Lagrangian parameter λ to achieve the target rate, was implemented and tested. In our experiments, the average number of iteration needed was 4. It appeared to be good enough for our needs, so no other approach was implemented.

Fig. 4.2 and Fig. 4.3 show that addition of MAP and MSE to the PSNR provided by using Lagrangian optimization is very small. $\mathtt{Lag}_{\mathtt{mse-linear}}$ adds from 0.05 to 0.1dB to the base increase of 0.6-0.9 dB that $\mathtt{Lag}_{\mathtt{none}}$ provides over $\mathtt{Sim}_{\mathtt{mse-linear}}$.

More examples about relative performance of different MAP and MSE methods can be found in Appendix B.

For Lagrangian optimization, MSE and MAP methods give results different from those we got for Simple transcoding. Fig. B.1 shows that MSE based on Linear interpolation is the best till a much higher rate reduction factor, and the advantage of $\mathtt{lag}_{\mathtt{mse-exp-var}}$ at factors close to 4 is very small. $\mathtt{Lag}_{\mathtt{mse-linear}}$ is above $\mathtt{Lag}_{\mathtt{none}}$ by 0.1dB for 3Mbps and by 0.05dB for 1Mbps. In Fig. B.2 and Fig. B.3 Linear interpolation based MSE approach is always the best one, but it gains only from 0.1 to 0.05 dB, depending on particular sequence.

MAP based on Linear interpolation pdf estimation still becomes better for lower rates, as we can see in Fig. B.4, Fig. B.5 and Fig. B.6, but it does not succeed to be better than MSE. $\texttt{Lag}_{\texttt{map-linear}}$ is about 0.1 dB above $\texttt{Lag}_{\texttt{none}}$ is the best case (for 1Mbps). So for most of the cases $\texttt{Lag}_{\texttt{mse-linear}}$ is the best solution.

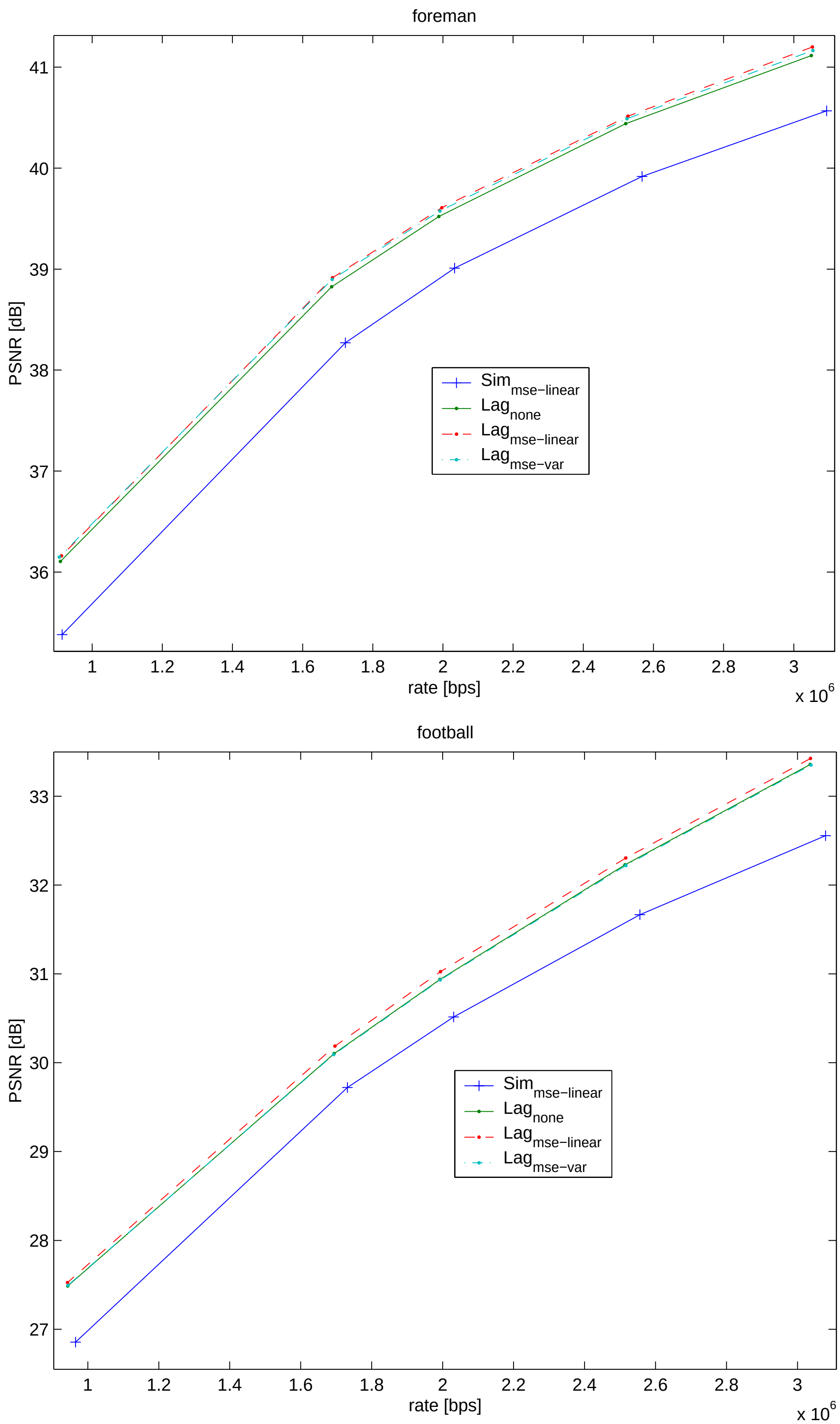


Figure 4.2: Comparison of Lagrangian vs. Simple transcoding for different sequences.

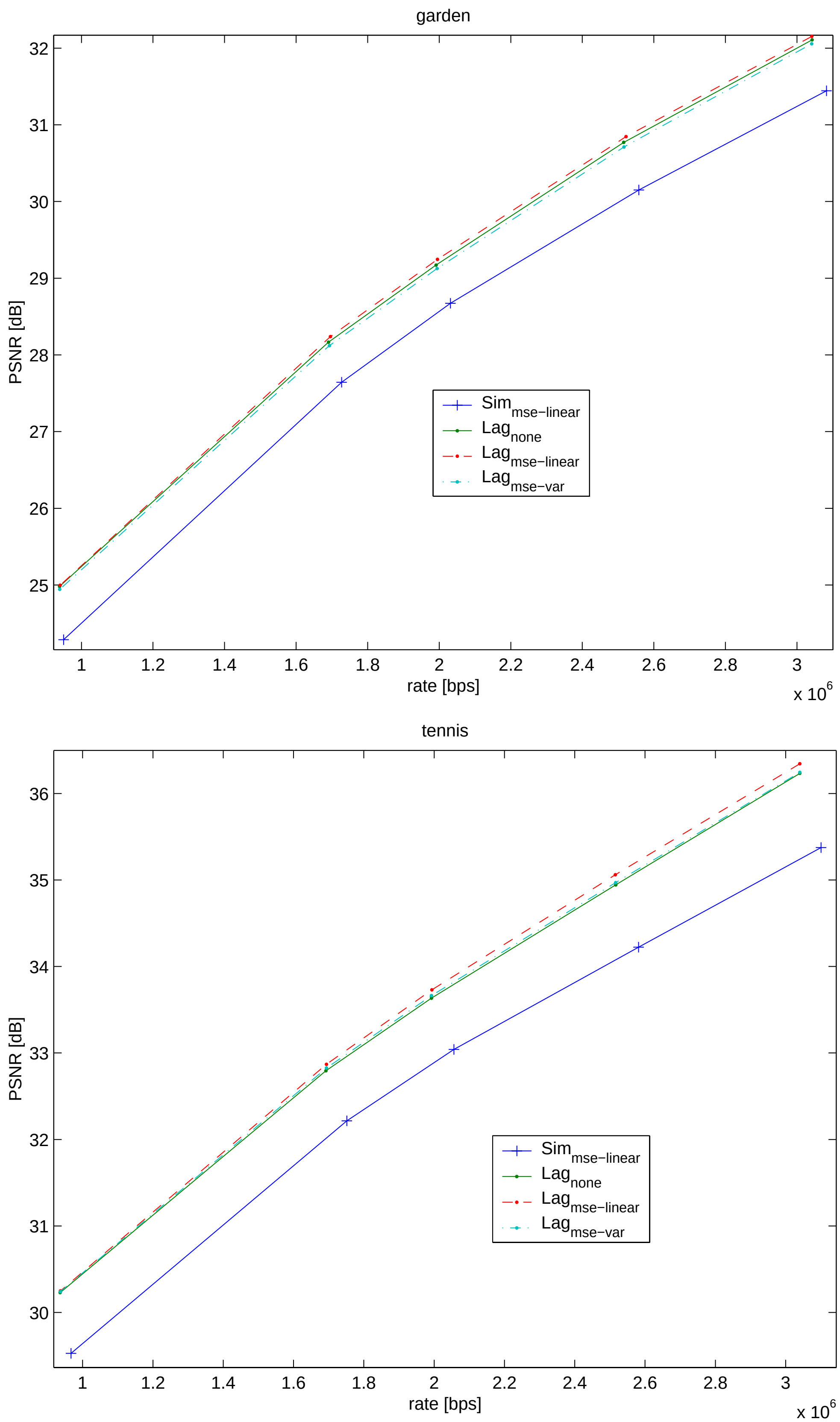


Figure 4.3: Comparison of Lagrangian vs. Simple transcoding for different sequences.

Chapter 5

Extended Lagrangian optimization

In this chapter we propose a novel extension of the Lagrangian optimization method by applying modifications to the quantized DCT indices. The proposed method outperforms (in terms of PSNR) all currently known requantization-based transrating approaches. To reduce the algorithm's complexity, we provide a low complexity trellis-based optimization scheme, and discuss other complexity reduction means as well.

In section 5.1 we introduce the quantized DCT indices modification method and define an extended Lagrangian minimization problem. Section 5.2 presents an effective solution based on a modification of the Viterbi trellis-search algorithm. Complexity considerations and means for its reduction are discussed in section 5.3. Experimental results are presented and discussed in section 5.4.

5.1 Quantized DCT indices modification

The idea of modifying the levels of quantized DCT coefficients, before applying VLC, for bit-rate reduction was proposed in [35, 17]. However, [35] discusses only methods for excluding AC coefficients in I-frames, and [17] considers only discarding several last non-zero coefficients in the zig-zag scan. We propose to extend the Lagrangian optimization presented in the previous section to include the possible modification of the values of all quantized DCT coefficients in an efficient way. The suggested optimization procedure aims at choosing quantized AC coefficient vectors values, as well as optimal quantization step sizes, that will provide a bit-rate that is as close as possible to the desired one with minimal distortion possible for the same rate.

Direct encoding of requantized coefficients using a given fixed run-level coding table does not necessarily provide the minimum possible distortion for a given total rate. It is possible to reduce the bit-rate needed for encoding by changing the values of the quantized vectors before run-level coding. To preserve the best possible quality at a given bit-rate, the selection algorithm uses a distortion penalty, caused by selecting reconstructed values away from the optimal one. An improvement, compared to selecting optimal quantization step sizes only, is expected due to the following reasons. For a particular quantization step size, it is possible sometimes to reduce the distortion if the cascaded quantization value can be changed to a value that provides a better approximation to the optimal reconstructed value. It is also possible to reduce the total bit-rate by breaking VLC pairs with long runs into several smaller ones that give a smaller total bit-rate than that of the initial pair. Of course, it increases the distortion most of the time. The goal is to reduce the bit-rate needed to encode some of the coefficients so that the distortion in other coefficients can be reduced.

Table 5.1 gives an example of possible quantized run-level pair changes. Let 0000004 be the quantized run-level pair to be encoded by the MPEG-2 VLC table. Suppose that we can choose if to encode it as is (i.e. without introducing any additional distortion D); to enter a 1 instead of the last 0, or to change the value 4 into

1. D here is the standard Squared Error. The number of bits needed to encode every sequence is defined by the *vlc* entry of the table. Because changing 0 to 1 splits the original run-level pair into two, for this case the *total* column in the table summarizes the cost of each pair to be used for making the final decision. Cost function $J(\lambda)$ for each case is shown for $\lambda = 0, 1, 3$. Minimal cost for every λ is in bold face. The Lagrangian multiplier λ defines the relative importance of rate vs. distortion introduced by changing the initial sequence. For small λ there is no reason to introduce distortion, as shown for $\lambda = 0$. If the weight of distortion and rate in the total cost function J is the same ($\lambda = 1$), it is better to split the run-level pair. In case λ is big enough, like $\lambda = 3$, a bigger distortion can be afforded to reduce the bit-rate.

The suggested optimization procedure aims at selecting quantized DCT coefficient values, as well as optimal quantization step sizes, that will provide a bit-rate that is as close as possible to the desired one, with minimal distortion.

As in the optimization problem stated in section 4.2, we may select a different quantization step-size for each MB, but here we also allow changing the quantized DCT coefficient value *by modifying its level (quantization index) value* after a particular quantization step-size has been applied. The minimization problem stated in

			0000014		
sequence	0000004	0000001	000001	4	*total*
run	6	6	5	0	-
level	4	1	1	4	-
vlc	24	8	7	6	13
$D = J(\lambda = 0)$	**0**	9	1	0	1
$J(\lambda = 1)$	24	17	8	6	**14**
$J(\lambda = 3)$	72	**35**	22	18	40

Table 5.1: Optimal quantized vector change for different values of λ

Eq. (4.1) remains the same, but now (4.2) is replaced by:

$$D = \sum_{k=1}^{N} d_k(q_k, \mathbf{v}_k),\ R = \sum_{k=1}^{N} r_k(q_k, \mathbf{v}_k), \tag{5.1}$$

where $\mathbf{v}_k$ denotes the index vector obtained by rounding the result of dividing the value of each DCT coefficient by the quantization step-size. All other parameters remain the same as in Eq. (4.2). Note that this formulation can be also applied to the initial encoding of the original data. The problem is still separable at the MB level - like in (4.4). But now, for every q_k, an additional minimization over all possible $\mathbf{v}_k$ values must be performed.

The same problem statement can be used for direct encoding of the original image, except for the fact that in transcoding we use an approximation of the original signal from previously encoded data. Distortion is usually measured by Squared Error from the source video sequence in the pixel domain, but it is possible to use linearity of the DCT and calculate the distortion directly from the DCT coefficients, without performing full decoding. Again, for every q_k, an additional minimization over all possible $\mathbf{v}_k$ values must be performed. Thus, for the k-th MB, Eq. (4.4) takes the form of:

$$J_k(\lambda) = \min_{q_k} \min_{\mathbf{v}_k} \{d_k(q_k, \mathbf{v}_k) + \lambda r_k(\mathbf{v}_k)\}, \tag{5.2}$$

and the set $\{\mathbf{v}_k(\lambda_s)\}_{k=1}^{N}$ is the optimal set of quantized vectors that provides the minimum distortion for a given total rate constraint, R_T. Here r_k depends explicitly only on the index vector $\mathbf{v}_k$, and not on the initial DCT values. Since the algorithm modifies $\mathbf{v}_k$ directly, and not via q_k, r_k is not a function of q_k any more. An efficient solution for the stated problem is proposed in the next section.

5.2 Trellis-based implementation

In this section a Trellis-based implementation of the above Lagrangian optimization procedure is discussed.

Let's define each location in the zig-zag scanned quantized DCT coefficients vector as a different stage in a trellis (Fig. 5.1). The cost value of a path is the sum of the costs of run-level pairs defined by this path. The optimal path up to a particular stage is the path that has the minimal cost value over all possible paths ending at that stage. The essence of a trellis-based algorithm is the fact that minimization of the cost value at each state of the current stage is the minimization of the sum of the current stage local-cost at each state and the minimal path cost already calculated at the previous stages of the trellis. It turns out that for the current problem, where different run-lengths need to be considered, the conventional trellis needs to be modified, so that every decision in a given stage does depend on previous stages, but luckily only on a single, already determined, state in each previous stage, as described below.

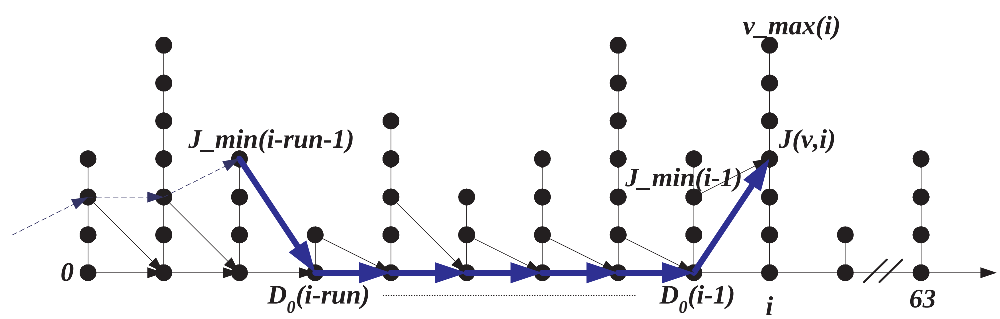


Figure 5.1: Trellis diagram for i-th AC coefficient in zig-zag ordered quantized coefficients vector.

Fig. 5.1 shows how the cost function is evaluated for a particular stage in the trellis. For trellis stage i (corresponding to the i-th coefficient) we have states from zero to $v\_max(i)$. $v\_max(i)$ is determined by multiplying the original index value by the initial quantization step-size, and dividing by the new one, followed by rounding upwards. In general, every possible v, $0 < v \leq v\_max(i)$, should be examined to see if it minimizes the total cost function $J(v, i)$ in Eq. (5.3) below. This cost depends not only on the value of v, but also on the number of zeros, i.e., the *run* leading to it, which defines the run-level pair for the VLC:

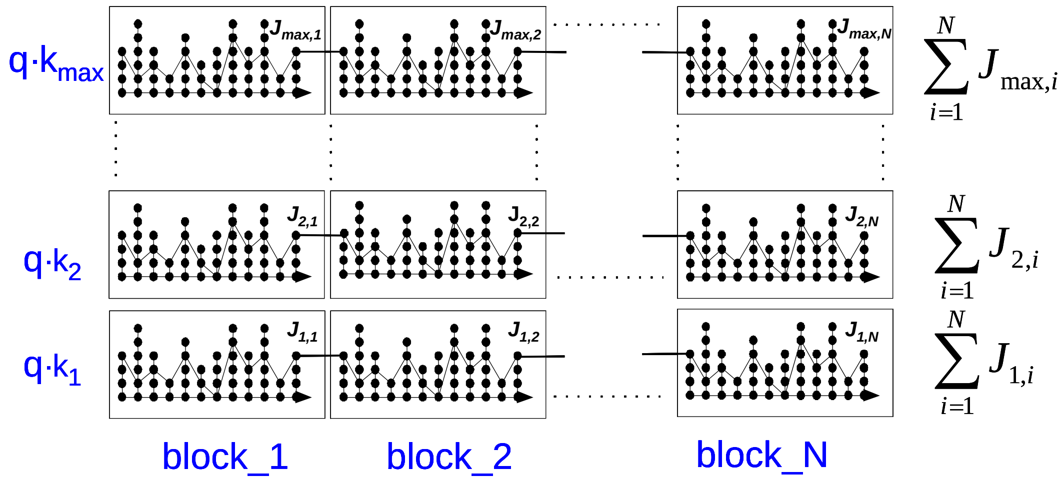
$$
\begin{aligned}
J(v,i) \qquad &= \min_{run}\{J\_min(i - run - 1) + \\
&+ \sum_{j=i-run}^{i-1} D_0(j) + \lambda R(run, v) + D(v, i)\}
\end{aligned}
\tag{5.3}
$$

where,

$J\_min(i - run - 1)$	- the cost of the minimal path up to the stage $(i - run - 1)$
$D_0(j)$	- the distortion caused by zeroing the j-th DCT coefficient
$D(v, i)$	- the distortion introduced by choosing v to be the value of the i-th quantized coefficient
$R(run, v)$	- the number of bits needed to encode the run-level pair (run, v) using the VLC

The dotted thin line on the left of Fig. 5.1 shows the minimal path till stage $i - run - 1$, which has the minimal cost $J\_min(i - run - 1)$. Thin arrows connect the last values of optimal paths in the previous stages to zero; or, in the case of stage $i - 1$, directly to the value v in stage i that is being examined. Different run-lengths need to be examined, but for a particular run the optimal state in the preceding stage is already known. The heavy line indicates the optimal path for the particular value of v. To determine the optimal value of v, the minimum over all its possible values of v has to be found:

$$
J\_\min(i) = \min_{0 < v \leq v\_max(i)} J(v, i) \tag{5.4}
$$



$$\sum_{i=1}^{N} J_{\max,i}$$

$$\sum_{i=1}^{N} J_{2,i}$$

$$\sum_{i=1}^{N} J_{1,i}$$

Figure 5.2: Trellis algorithm on Macroblock level.

Fig. 5.2 shows the schematic view of the proposed trellis diagram to be used in the optimization algorithm on a Macro-block level. The trellis is built for every requantization step-size, that is presented as rows in Fig. 5.2. Corresponding step-sizes are denoted by multiples of the initial quantization step-size q. Trellis in each row is divided into sub-trellises on a block level. For a particular block and $q \cdot k_i$ step-size the range of quantized AC coefficient values is built (dotted columns). The optimal path (thin lines) throughout a block is calculated for a particular λ as was described above. For fixed λ and $q \cdot k_i$, the cost function $J_{i,j}$ can be calculated independently for each block j in the particular MB. At the end of each block the best quantized vector is chosen in every row. For the next block the optimal path pointer of the previous block trellis is provided, as shown by a block connecting line. At the end of each macro-block the optimal quantization step size is chosen:

$$q_{k,opt}(\lambda) = \underset{\{q \cdot k_i\}_{i=1}^{max}}{\arg\min} \sum_{j=1}^{N} J_{i,j}(\lambda, q \cdot k_i) \tag{5.5}$$

Yet, even when the above trellis is used, the number of calculations needed to perform the optimization is rather high. Hence, in the next section we consider ways to speed up the algorithm.

5.3 Complexity Considerations

The proposed method needs, in principle, many iterations over a large number of parameter values because:

1. The number of examined runs for every *level* in a particular stage increases with the index value of the DCT coefficient being processed.

2. All stages, till the stage corresponding to the last non-zero coefficient, in every block, need to be examined.

3. A separate trellis has to be constructed for every requantization step-size that we wish to examine.

4. There are several values of v at each stage that need to be examined.

5. Several values of λ need to be tried (in a directed way) before the total rate will match the constraint

Note, however, that while the number of index values (levels) to be examined at each stage seems to be large at first sight, it is no so in reality. This is because the mean value of the AC coefficients is typically in the range of 30-50. Hence, for example, if the initial quantization step-size is 6, then even on the finest scale there are on average only about 5÷10 values to choose from. When the quantization step is increased, we reach a single value very quickly.

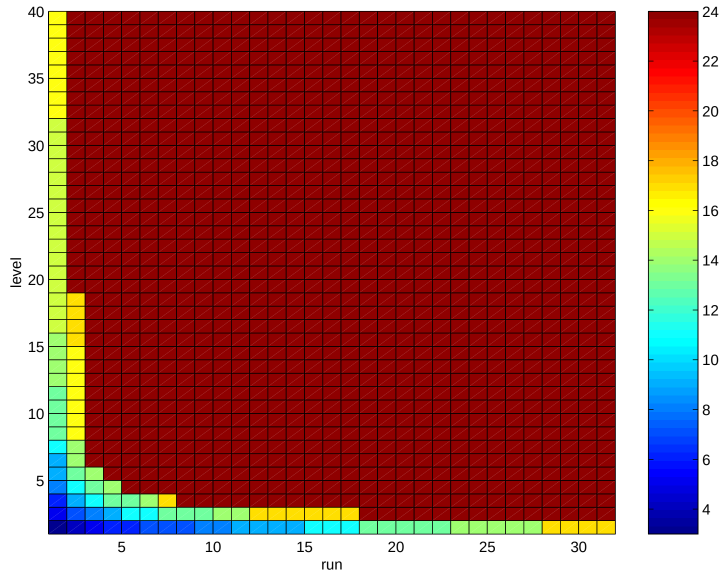


Figure 5.3: one of two possible MPEG-2 VLC tables.

As for quantization step-sizes, there were a number of works that propose to restrict the range of requantization step-sizes based on the initial MB quantization step-size [27, 28, 23].The most recent [23] shows that for open-loop transcoding Lagrangian optimization [4] can be restricted to check even and odd multiples of the initial quantization step-size for I and P-B frames, respectively. At those ratios many quantized coefficients are zeroed out after rounding. The optimal multiples of the initial quantization step-size are changing according to different quantizers used for I and P,B-frames, as shown in Fig. 3.5. The results reported for the closed-loop scheme were very close to those of full Lagrangian search. As for searching over different values of λ, applying a simple bi-section search, as in [5], requires on average about 3 iterations only.

5.3.1 Complexity Reduction Means

As mentioned above, the number of examined runs for every *level* in a particular stage increases with the index value of DCT coefficient being processed. Let's have a look on one of the MPEG-2 VLC tables presented on Fig. 5.3. We can observe that, practically, the number of level values that should be considered for obtaining a rate reduction is actually not that large. Moreover, if we consider choosing a run for a particular level v, the number of options to examine - $run_{max}(v)$, before getting to the maximum no. of bits in the VLC table, $R_{max} = 24$, is very small for most levels (the exceptions are levels 1 and 2, for which there are 31 and 16 possible runs, respectively).Thus, Eq. 5.3 can be rewritten as follows:

$$J(v,i) = \min\left\{\left(J^{opt}(v,i), J_{opt}(run_{max}(v),i) + \lambda R_{max}\right)\right\} + D(v,i) \tag{5.6}$$

where,

$$J^{opt}(v,i) = \min_{run < run_{max}(v)} \{J\_min(i-run-1) + \sum_{j=i-run}^{i-1} D_0(j) + \lambda R(run,v)\} \tag{5.7}$$

is the part that depends on exact (*run, level*) values chosen and thus needs to be evaluated for every *run*, and

$$\begin{aligned} J_{opt}(run_{max}(v),i) &= \min_{run > run_{max}(v)} \{J\_min(i-run-1) + \sum_{j=i-run}^{i-2} D_0(j)\} + D_0(i-1) \\ &= J_{opt}(run_{max}(v)-1, i-1) + D_0(i-1) \end{aligned} \tag{5.8}$$

is the part we know prom previous calculations.

To calculate $J^{opt}(v,i)$, $run_{max}(v)$ iterations are needed, while $run_{max}(v)$ is usually a small number. $J_{opt}(run_{max}(v) - 1, i - 1)$ is known from the last stage, so no search needs to be done to find $J_{opt}(run_{max}(v), i)$. So, to get $J(v, i)$, only $run_{max}(v) + 1$ calculations need to be performed. Using Eq. (5.6), (5.7), (5.8) instead of Eq. (5.3) *does not affect the optimality of the solution*, but reduces the number of calculations needed by up to 40% - in our simulations.

Another simplification that is proposed here results in a sub-optimal solution. As was mentioned above, sometimes it may be useful to split a run-level pair into two smaller ones. It is reasonable to insert the minimum possible non-zero value because it provides minimum distortion, and also possibly the minimum rate because the run-level pair bit cost is an increasing function of level. It is difficult to say which will be the optimal one because of the usage of a quantization matrix that changes the distortion weight for every AC coefficient. In case someone wants to use a distortion metric different from MSE, things will become even more complicated. If no splitting is allowed, the trellis needs to go through stages defined by non-zero coefficients only. In common MPEG-2 encoded block there are in about 70-90 % of zeros, so the computational complexity reduction is very pronounced.

5.4 Experimental results

As the proposed extension to Lagrangian optimization minimizes the error over quantized DCT indices, the MAP approach, which minimizes the error probability, is not applicable in this case. Still, MSE estimation can be used here. MSE that is based on Linear interpolation ($\mathtt{Tr}_{\mathtt{mse-linear}}$) becomes the best method for all the sequences. As in sections 3.5 and 4.3, the gain in using the MSE gets smaller for higher compression factors. For FOREMAN sequence, $\mathtt{Tr}_{\mathtt{mse-linear}}$ is from 0.02 to 0.05 dB above the

$\texttt{Tr}_{\texttt{none}}$, while for GARDEN and TENNIS sequences the revenue is from 0.05 to 0.1 dB.

For the sub-optimal Trellis search, the gain in using the MSE is slightly bigger - up to 0.15 dB for 3Mbps.

Fig. 5.4 and Fig. 5.5 summarize the best results of the previous chapters. We compare the performance of original encoding (Enc) and cascaded decoding and encoding (Re) using a standard TM5 encoder, with the Simple complexity-based transrating (Sim), Lagrangian (Lag), Full Extended Lagrangian optimization using a Trellis diagram (Tr) and the sub-optimal Extended Lagrangian optimization (Trs). As MSE using Linear interpolation for pdf estimation is the best method in most of the cases (which is also true for Simple transrating, see sec. 3.5), only Linear interpolation MSE approaches are compared.

In our simulations, the proposed Extension of Lagrangian Optimization by DCT coefficient index modification is always the best (in terms of PSNR), and it increases significantly the improvement provided by the previous Lagrangian optimization scheme. For the FOREMAN sequence, $\texttt{Lag}_{\texttt{mse-linear}}$ increases from 0.25 to 0.7 dB above the PSNR of the originally encoded sequence, while $\texttt{Tr}_{\texttt{mse-linear}}$ gets from 0.6 to 1 dB above it. For other sequences $\texttt{Tr}_{\texttt{mse-linear}}$ can double the gain of $\texttt{Lag}_{\texttt{mse-linear}}$. For the TENNIS sequence, for rates below 1.6 Mbps, regular Lagrangian optimization can't become better than Enc, while $\texttt{Tr}_{\texttt{mse-linear}}$ is about 0.4 dB above it. At the same time the performance reduction by using the proposed sub-optimal Trellis search is very small for all sequences (less than 0.05 dB), except TENNIS, where it reached 0.09 dB.

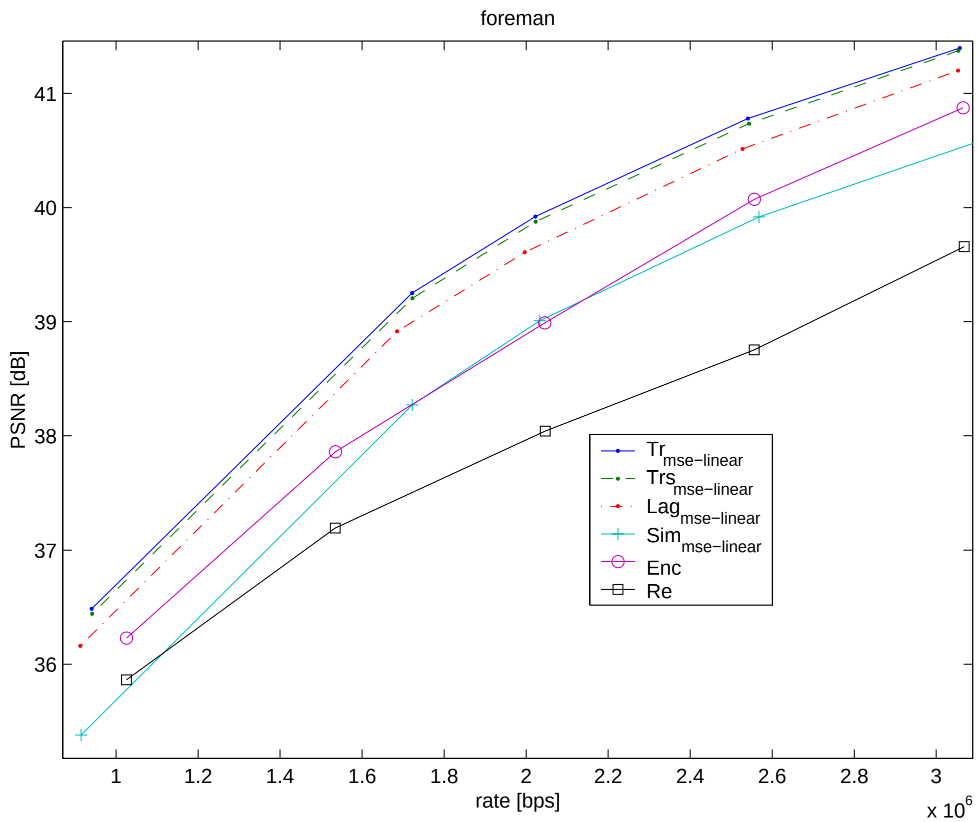
Additional examples can be found in Appendix C.

Another issue we wish to present here is computational complexity. In the table below we summarize the running time information achieved by using Microsoft Visual Studio profiler under Windows. To use the profiler, compiler code optimization was disabled.

method	run-time(msec)	ratio
Simple	50	0.15
Run-time optimized Lagrangian	330	1
Extended Lagrangian	7150	21.7
Extended Lagrangian with optimized search	2600	7.9
Sub-Optimal Extended Lagrangian	1100	3.3

Table 5.2: Run-time of different transrating approaches. For "Simple" method run-time is measured over one frame. For other algorithms the time is for one iteration with particular λ.

The proposed Extended Lagrangian optimization is 8 times slower than the fastest regular Lagrangian method. However, using the proposed sub-optimal solution reduces only slightly the PSNR, and is only 3.3 times slower than the regular Lagrangian optimization.



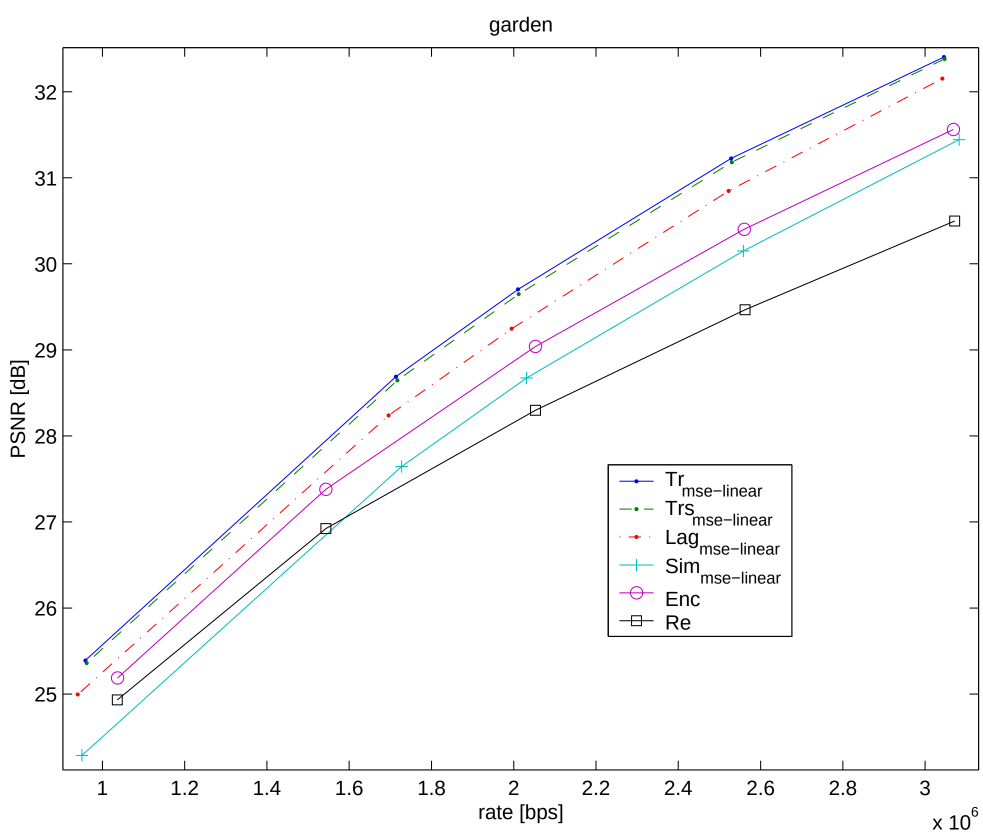


Figure 5.4: Best MSE transrating for all methods for different sequences.

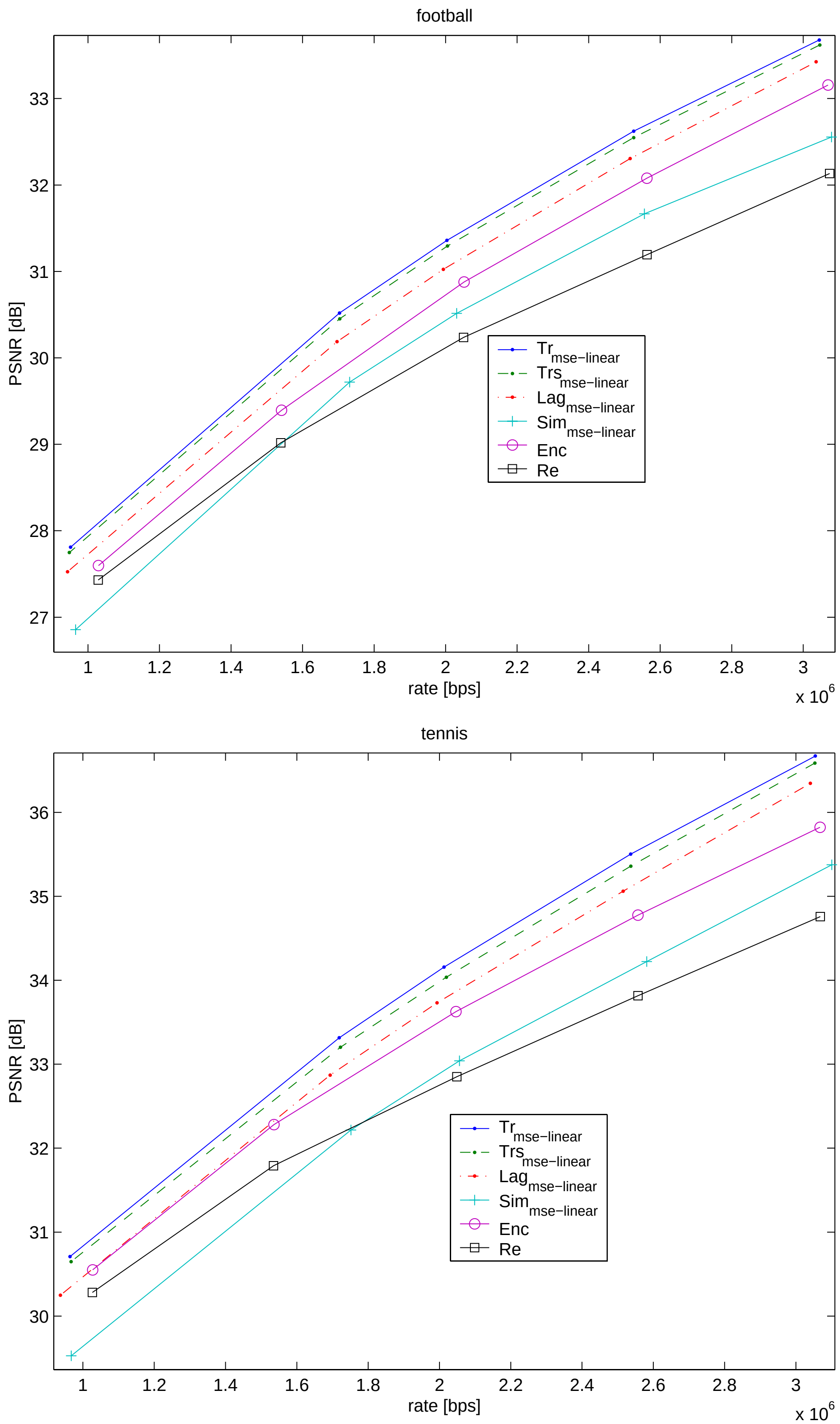


Figure 5.5: Best MSE transrating for all methods for different sequences.

Chapter 6

HVS-based segmentation and tracking

In video production and editing, there is common agreement that not all of the video picture parts are of the same importance to a human observer. There is no need to encode the entire scene at the same quality. The solution is to segment the picture and to process it according to its perceptual importance. There is a lot of effort to define HVS models and to implement encoding, editing and display based on those models. Most of the works are done in the pixel domain framework, in spite of the fact that HVS models show the importance of neighboring pixel interaction, both in space and time, for perceptual quality of particular area of video sequence. Spatial correlation of pixels is utilized by frequency transformations used by modern encoding systems like Fourier Transform (FT), DCT, Wavelet Transform (WT), and others. Pixel based methods looks like the most popular for encoding because it is relatively easy to extract boundary and object information. At the same time, transform information of already compressed data for video editing and post-processing before displaying the decoded video seems to be a better solution. The following section aims at video segmentation for transrating applications. Transrating input is compressed-domain data, and so is the output. More precisely, the current research deals with MPEG-2 transrating, i.e. reducing the bitrate of already block-based DCT encoded video. Section 6.2 describes the segmentation scheme we used in this work for transrating. Results of the proposed scheme are shown in section 6.3.

6.1 Existing Compressed Domain Segmentation Methods

Following [36], features used by compression domain segmentation schemes are:

1. DC value of DCT, which is the mean luminance value of the block, known as μ in pixel-based methods.

2. Sum of squared AC coefficients, which corresponds to the variance in pixel domain.

3. Sums of amplitudes of AC coefficients in the first row and first column of the DCT coefficient matrix was proposed as a measure for vertical and horizontal edge presence in a block by [37]. More complex DCT domain edge detection method are proposed in [38].

4. Color: DC values of Y, Cb and Cr color component blocks can be used to form a block's color vector[39].

5. Motion Vectors (MV) presented in Inter-coded pictures and blocks.

6. MB bit-count information multiplied by quantization step size can provide information about coding complexity of the particular MB that shows how hard it is to reduce the MB bit-count by increasing the quantization step-size.

7. MB coding type must be taken into account while making decisions about specific block classification.

Segmentation approaches can be divided into two main groups:

1. Local-properties-based methods. These methods assume that what is important for a human observer looking on a particular region of a picture are the local region properties themselves, with some minor correlation to the adjacent areas features, like mean luminance.

2. Object-based methods. These methods presume that a human observer identifies different objects in a video sequence, and what is important to him is the better quality of the important video objects in the scene.

6.1.1 Methods based on local properties

These methods are based on extracting local features from compressed domain data, like blockiness [40] or boundaries [38], which can be used to define the perceptual importance of a particular image block. It is also important to remind here works that aim to define perceptual quality based on the DCT domain information. Some of those works still need the original picture for comparison [41], which is not the case we have in transrating.

In [42] it is proposed to calculate perceptual activity of every block based on the DCT coefficients weighted by a correlation matrix computed in a bigger neighboring area centered at the same block. The ratio of a particular block and neighboring areas DC values is supposed to take into account Weber's law. This ratio is appropriately combined with the sum of absolute values of weighted DCT AC coefficients to measure the perceptual activity of that block.

Considering implementations for transrating, it turns out that there are not so many works published that are based on local properties as there are about object-oriented methods. The methods utilized are also much simpler.

In [43] a Hybrid Transrating method that switches between requantization, resizing and frame skipping was proposed. The decision is made at frame level, so the measures proposed were Motion Activity, which is the average magnitude of the motion vectors of all the macroblocks in each frame, and Spatial Activity, which is defined as the mean quantization step-size, or the actual number of bits used to encode the frame if the first the mean quantization step-size exceeds its maximum value.

Another frame-based MPEG properties extraction that aims at transrating was proposed in [44]. Quantization step-size is called Region Perceptibility and the number of zero DCT coefficients in the block is defined as Spatial Complexity. As Temporal Similarity features, it is proposed to gather MB coding type, the direction of MVs (forward, backward or bi-directional), prediction type (frame or field), and the number of actual MVs used for a particular MB (from one up 4 in bi-directional field predicted MB). The work deals mainly with MPEG-2 to MPEG-4 transrating and does not target requantization, i.e. transrating problem. So the major aspect analyzed was which properties can be re-used, and which has to be modified and how to meet MPEG-4 standard specifications.

Some encoding techniques can be adapted to transrating. It is mentioned in [45] that objectionable artifacts that occur when pictures are encoded at low bit rates are: *blockiness, blurriness, ringing* and *color bleeding*. Blockiness is related to high quantization of smooth regions, while ringing and color bleeding occur at edges on flat background where high frequencies are poorly quantized. Blurriness is the result of loss of spatial detail in moderate- and high-detailed regions. To avoid producing those artifacts, the MBs are classified into either homogeneous or edge-including, based on a comparison of MB block's variances. Homogeneous MBs are further sorted into several classes from flat to coarse-textured. At the same time edge-including MBs are arranged into sub-classes as weak edge, normal edge, strong edge and structured edge. A similar approach can be implemented for transrating, with a further bit-rate allocation that is proportional to each class perceptual importance, as it is done in encoding.

6.1.2 Object oriented Methods

There are many efforts to perform object segmentation of a video stream. Most of the works were motivated by the MPEG-4 standard that enables Video Object Planes (VOP) encoding. To prepare tools for MPEG-2 to MPEG-4 conversion, many authors propose segmentation algorithms working in the compressed domain.

Several researches try to segment the video on the basis of MV information. MVs can be accumulated over time to build dense motion information for every pixel in the frame and to use it for pixel-domain segmentation. Such segmentation can maximize the probability that all pixels in a particular segment share the same affine transformation [46]. The pooling of MVs from multiple frames is also used for building a more robust data set for clustering [47]. To withstand MV estimation errors, the data set includes all MBs with MV close enough to currently estimated affine models, and after that the data is re-clustered, and the affine models are updated. Segmentation tracking needed in transrating is usually exploited in background freezing approaches that can be used for fixed camera scenes. If an initial block-based segmentation is known, the tracking can utilize MV information of encoded MPEG stream to detect what new blocks are covered by projections of the initial object MBs [48]. The MBs are divided into Active (predicted objects), Monitored (MBs that are close enough to the objects) and Inactive. A set of rules define how some objects MBs can be transferred either to Monitored or Inactive clusters, and how MBs from Active or Monitored partitions can be transferred to an Active set.

DCT coefficients similarity can be defined and used for image segmentation in many different ways. For example, it is possible to calculate the DCT of a block centered at every pixel in the picture, and by PCA to define the number of dominant regions in the scene. By using appropriate filters it is possible to fit the DCT coefficients to frequency characteristics of HVS, and so get perceptually adapted segmentation after K-means clustering on what is called Situational DCT Descriptors[49].

The combination of DCT and MV information for moving objects extraction was also tested in a number of works. Region Growing methods often use a spatial feature vector that consists of mean block luminance, vertical edge, horizontal edge and texture (see section 6.1 features 1,2 and 4). Following segmentation, the region is set to be a moving region if more than a half of its MBs have non-zero MVs [50]. This approach is sensitive to MV estimation errors, so the possibility of false motion has to be minimized by comparing coded motion residual with no motion residual, which has to be calculated additionally by the proposed scheme. A somewhat more complex classification of false motion blocks was proposed in [51]. It checks if blocks adjacent to a particular block in the direction of its motion are also dynamic. A possible combination is assigned using the ratio of probabilities of the realization of this combination assuming "real moving block" to that assuming "false moving block".

Another way to segment Intra frames is to employ the Watershed method [52], which is widely used in the pixel domain, to a simplified image consisting of the DC values only or DC plus 2 AC coefficients picture [53]. The Watershed method is used to produce closed smooth areas coverage of the initial image, based on a gradient image. One of the method drawbacks is that it usually over-segments the image due to local noise in the picture. To avoid over-segmentation, leveling that uses as a marker the DC image and the DC+2AC image as reference is employed. Markers mark some of the initial zones as areas of interest, and the segmentation is constructed starting from these zones, so better segmentation can be achieved. The largest flat zones in the resulting image are chosen as markers for a final watershed segmentation, which is based on morphological gradient of the leveled image. To find the moving regions, the MVs of neighboring Inter frames are summed for every MB, and the Manhattan distance is calculated on resulting MVs to represent the displacement, which is further uniformly quantized into eight levels. This information can be used for final segmentation, or for global motion estimation of the scene.

A more complex spatiotemporal segmentation scheme is proposed in [54]. The initial segmentation is generated by Sequential Leader Clustering of vectors combined from DC of all three image color components and AC energy information. Sequential Leader Clustering algorithm uses pre-defined threshold and distance measure to decide if new point has to be added to the closest of the existing clusters, or, if the distance is above the threshold, has to start a new cluster. After a number of clusters is defined, adaptive K-mean clustering is applied to adjust the initial segmentation. The small regions are merged with their neighbors using luminance and AC energy distances. Entropy values of ac energy for every region is calculated, and modeled as a Gaussian distribution with mean μ and standard deviation σ, over all regions in the picture. Then, spatial similarity is calculated for every pair of adjacent blocks based on assumption that entropy difference must be of zero mean and of variance $\sqrt{2}\sigma$. The temporal similarity is derived based on the Kolmogorov-Smirnov hypothesis test of the distribution of the temporal gradient. Kolmogorov-Smirnov statistic is defined as the maximum value of the absolute difference between two cumulative distribution functions. Temporal gradient is calculated by applying a 3-D Sobel filter. It is not described in the letter [54] on which data in compressed domain, if at all, it was applied. Finally, two similarity measures, one for merging spatiotemporally similar regions (a modification of similarity measure proposed in [55]), and one for merging regions with high average temporal change within the region, are calculated. After Region Adjacency Graph (RAG) merging based on those similarity values, the regions are classified as foreground or background based on average temporal change of regions.

Another approach for clustering is Maximum Entropy Fuzzy Clustering (MEFC) in the compressed domain. The idea of fuzzy clustering is to iteratively calculate the region centers and to update each image sample membership based on some distance measure. In MEFC the distance from a pixel to every cluster is defined by the squared difference of the pixel value from each cluster center value that is weighted by a membership function. The membership function derivation is based on the principle of maximum entropy to yield a least biased clustering. It is possible to use MEFC on DC coefficients, and further to refine the membership of blocks surrounding the current object to be either foreground or background based on a Maximum A Posteriori (MAP) approach applied to those block's AC coefficients [56]. For video it was proposed to use MEFC for segmentation of MV data projections on I-frame from two P-frames surrounding the I-frame. To achieve finer VO's boundary segmentation, MEFC is further applied to DC coefficients of I-frames [57]. To avoid problems caused by non-perfect MV information, it is assumed to be of impulse noise nature, and is filtered by Noise Adaptive Soft-switching Median (NASM) filter [58]. The filtered MVs are then clustered by the MEFC algorithm. Validated regions are tracked using Kalman filtering, which update second order motion model. Segmented regions projection according to that motion data is used to make robust temporal segmentation, which tests motion homogeneity and overlapping area of the segmented masks, such that a recursive merging and splitting process could be performed in the temporal domain [59]. DC coefficients of three color components are then clustered into an optimum number of homogeneous regions, also by applying MEFC. To complete the classification of the small homogeneous regions that are usually produced by the proposed algorithm, MAP processing to classify those regions into either moving foreground or stationary background is proposed.

6.2 Proposed Scheme

In the current research, segmentation is used to provide different perceptual importance to different parts of the picture so as to match the MSE distortion in those areas to perceptual criteria. The algorithm described below uses the information available in the encoded stream, like AC coefficients and motion vectors, to perform a local-based segmentation. Each block in the picture gets a relative weight that carries information on the extent the rate can be reduced in the particular block during transrating.

The proposed segmentation scheme uses a somewhat simpler approach than [45]. It consists of two main units:

1. Encoded data segmentation unit, which partitions the picture into segments based on the coded AC coefficients.

2. Tracking unit, which utilizes motion information to track the segments variations over time.

The segmentation technique developed suits both open-loop and closed-loop transrating schemes. So, only the information available in the DCT domain and the motion vectors are used. As follows from the original goal of this scheme, the segmentation is done at a resolution of 8×8 blocks. It is possible to replace the proposed segmentation scheme by another one, and still utilize the tracking scheme to improve results over time.

6.2.1 AC-based segmentation

Following [45], we assume that the amount of information in smooth areas is small, hence so is its bit-budget. However, even a small reduction in their bit-rate typically results in a visible degradation in picture quality. On the other hand, textured regions in the picture demand many bits, but their rate can usually be significantly reduced, before a human observer will notice the difference. Boundaries (edges) are known to be the most perceptually important part of the scene, but in terms of rate vs. noticeable distortion, they fall, on average, between smooth areas and textures. Thus, the proposed scheme classifies blocks as being either smooth, textured or boundary. It is possible to define a textured block as block that is surrounded by other boundary blocks from almost all directions. It can either include boundary by itself, or not. The segmentation algorithm thus consists of the following steps:

1. Block activity evaluation. The sum of absolute values of AC coefficients can be a good measure for edge presence in a particular block [36]. DC and two first AC coefficients are omitted, as not to take into account slow luminance changes over flat areas.

2. Binarization. The block activity picture is converted into a binary picture that classifies each block as either a high-activity block or low-activity block. Binarization is done using adaptive thresholding. The Otsu method [60] uses an interclass variance criterion for bilevel thresholding. It is simple and usually gives very good results. This technique yields a threshold resulting in minimal intra-class variance and maximal inter-class variance.

3. Texture detection. Morphological operations are used to remove from the set of high-activity blocks the blocks that are close to relatively big low-activity areas. The size and the form of these areas are defined via a particular selection of structuring element. During these operations the edge between textured and flat areas is removed, as well as small isolated blocks and thin boundaries between smooth areas. Other operators are used to fill in small holes in the remaining set of high-activity blocks to form textured areas.

4. Forming segmented picture. At this point, textured regions are already formed. All the blocks that pass the threshold but were not included into the texture class are classified as boundaries. It includes isolated active blocks in low-activity areas, and boundaries between textured and smooth regions as well.

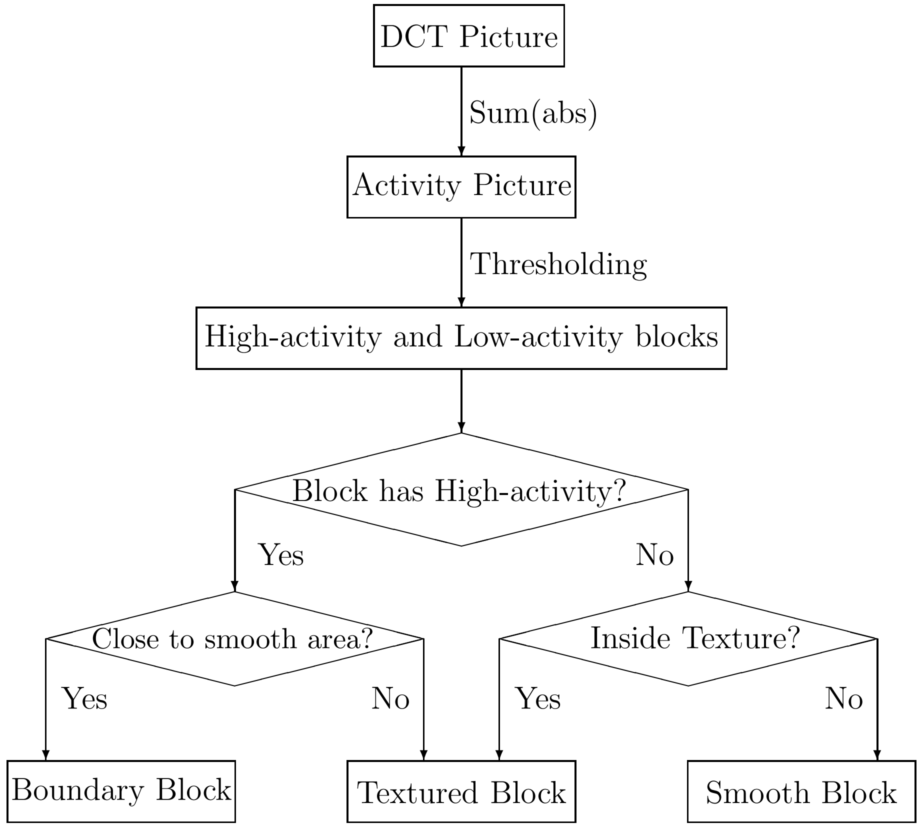


Figure 6.1: Flow-chart of the I-frame segmentation procedure. Morphological operations are used to decide if close to smooth area or inside texture.

Every low-activity block that is not included in the texture class is considered smooth.

Fig.6.1 presents flow-chart of the segmentation process. Decisions about low-activity blocks are taken after high-activity blocks are proceeded.

6.2.2 Tracking

MPEG video encoding uses differential encoding and motion estimation to reduce the amount of information to be encoded in each frame. While I-frames can provide reasonable segmentation results, P- and B-frames don't contain enough information by themselves. They do provide essential information about changes in the scene that can not be predicted from the reference frames, but it is impossible to get good segmentation of these frames without tracking.

The idea of the proposed scheme is to use motion vectors information to track the segmentation changes due to motion is the video sequence, and to update the tracking results by new information available in difference pictures. For each block type we define the grade that will be used to weight the distortion function during Lagrangian optimization procedure. This grade defines perceptual importance of a particular region type. More importance will be given to smooth areas than to boundaries and textures. Textured areas will be weighted so that the highest distortion can be introduced in them. In I-frames the grade will be given based on the block type only.

The segmentation scheme described above is performed at block level, while motion vectors have pixel or even sub-pixel resolution, so each tracked block usually depends on more than one reference block in reference picture. Because of partial covering of reference blocks, it is difficult to classify the block type exactly. The distortion level allowed in boundary blocks is between the distortion allowed in textured blocks, which is the highest one, and the distortion allowed in smooth areas, which is the lowest one.

There are several options to determine tracked block grade based on reference block classification. Consider the situation shown in Fig.6.2. Tracked block P has motion vector MV that points to reference area R. This area cover blocks $1, 2, 3, 4$, which are graded as r_1, r_2, r_3, r_4. The covered areas of each block are w_1, w_2, w_3, w_4, respectively. It is possible to produce a classification that is based on the maximally

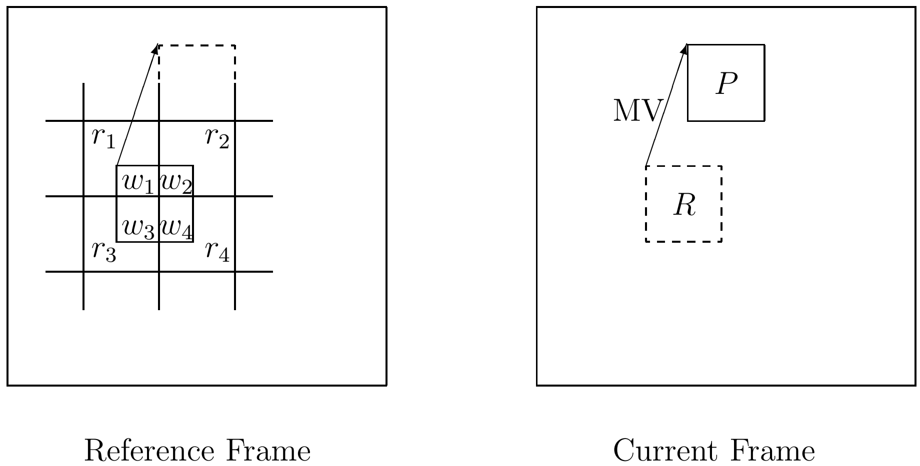


Figure 6.2: Block tracking over time. As tracked block P is in current frame, and reference area R is in reference frame, their projections to reference/current frames are shown by dotted squares.

covered reference block:

$$r_P = r_{argmax\{w_1,w_2,w_3,w_4\}}, \tag{6.1}$$

or to classify it to the class of the reference block that is the most distortion-sensitive of the four reference blocks:

$$r_P = \min\{r_1, r_2, r_3, r_4\}, \tag{6.2}$$

Another option is to use the sum of reference-blocks grades, weighted each by their covering area of reference blocks:

$$r_P = \sum_{i=1}^{4} w_i \times r_i, \tag{6.3}$$

We implemented the tracking as follows:

1. Reference-blocks determination. In the case of field-encoding, a B-field can

have up to four reference field-pictures. In every picture the reference area of a particular block consists of one to a maximum of 4 blocks. If field-encoding for blocks in a frame-coded picture is used, things are even more complicated. Reference block type and covered area used for constructing the reference have to be accurately tracked. I-MBs in P-,B-pictures have no reference, and this has to be taken into account on later stages.

2. Block classification. Based on reference segmentation pictures and block-cover information, tracked blocks are classified. The proposed scheme gives tracked blocks a grade according to its probability of being a particular type. The classification grades of all reference blocks are weighted according to reference area covering of each block and their sum is assigned to the tracked block.

3. New data segmentation. To account for changes that can't be fully tracked on the basis of information from previously segmented frames, the same segmentation scheme used for I-frames is applied. This segmentation can't provide a final segmentation, but can detect new boundaries and textures. The resulting boundary and texture segments are merged with the tracking results from step 2, based on the assumption that merging 'edge' and 'texture' types gives 'texture', but a new edge may appear in a previously smooth area.

6.2.3 Perceptual weighting

To take into account the different perceptual sensitivities at smooth, textured and boundary areas, we propose to modify the distortion cost function by multiplying the SE of each block by a weight function. After a number of experiments we found that reasonable weight-ratios for the different area types are 4:2:1 for smooth, boundary and texture, respectively. A larger weight gives more importance to a particular area, so less distortion will be introduced there.

Up to this point the proposed system uses the same features that are usually used for still images, i.e. no consideration was given to motion. The following enhancement utilizes temporal domain characteristics of the HVS when viewing video.

It is known that HVS is less sensitive to distortion in quickly changing areas. We decided to measure temporal changes in the video signal by means of the available motion vectors. We observed in experiments that motion up to 8 pixels in consecutive frames, in any direction, should not result in a modification in the initial weights, since the observer can track those changes easily. The faster the motion is, the change in the same region becomes larger, and thus there is less sense in preserving the same quality in every frame there. The weight modifier that we propose is:

$$W_{mv} = \frac{W}{f(\sqrt{MV_x^2 + MV_y^2})}\ ,\ f(x) = \begin{cases} 1 & if \quad x < 8 \\ 1 + k * (x - 8) & \quad x \geq 8 \end{cases} \tag{6.4}$$

where,

W_{mv}	- motion modified perceptual weight
W	- initial block weight determined by segmentation and tracking
$MV_{x/y}$	- motion vectors x/y components of current MB
k	- a constant (set to $k = 1/24$)

6.3 Simulation results

Evaluation of perceptual-based coding methods, as well as optimal parameter selection, is an ill-posed problem because there is still no standard *objective* video quality metric. The ITU-T Recommendation P.910 standardize methods for multimedia *subjective* quality assessment. Because the premise behind subjective assessment is the use of human observers to rank video sequences, it is impractical and impossible to use these methods for HVS model parameter adjustment, and it is very hard even for the evaluation of very limited sets of data.

In spite of the fact there is no standardization yet in this area, there are some Perceptual Quality Assessment (PQA) tools provided by different companies. We got an opportunity to run some tests of our scheme with a commercial Tektrnix/Sarnoff PQA200 tool, access to which was kindly provided by Optibase. The PQA200 measures a 2-second portion of a 5-second video test sequence. The measurement results in a single numeric value of picture quality called Picture Quality Rating (PQR). Utilizing a human vision system model, the PQA200 claims to analyze the three necessary dimensions for evaluation of dynamic and complex motion test sequences: spatial analysis, temporal analysis and full-color analysis.

For carrying out the tests, we got from Optibase three video sequences - MOBILE, FOOTBALL and CHEERS - in 720x480, 4:2:0 resolution. Those sequences were encoded at 6Mpbs and 4Mbps with a standard TM5 encoder. The 6Mbps video was transrated to 4Mbps using different transrating methods, described in previous chapters of the thesis, as well as Lagrangian optimization based transrating methods, which are perceptually weighted as it was proposed in this chapter. Evaluation of a particular video sequence took about 7 minutes. We also tried to evaluate transrating from 6Mbps to 3Mbps, but faced problems because the distortion introduced in the special marks used by PQA200 for synchronization, so only the results of transrating to 4Mbps will be shown here. The results are probably lower than could be achieved because our transrating system is working with *frame-encoded* MPEG-2 video, while

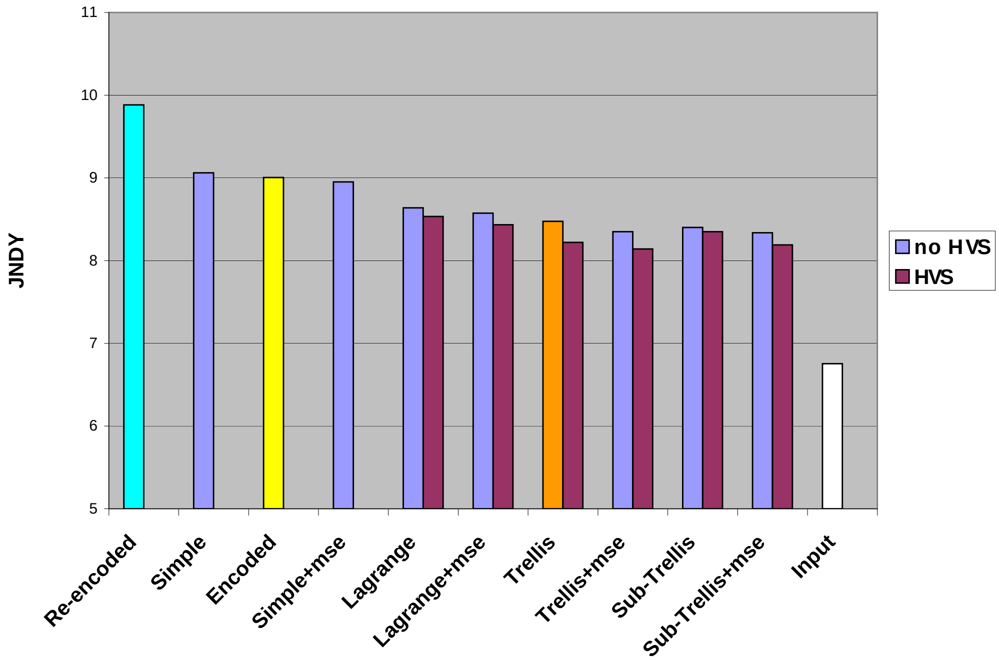


Figure 6.3: Perceptual Evaluation Results: FOOTBALL sequence transrated from 6Mbps to 4Mbps.

the sequences we got for evaluation are *interlaced* videos, which are better suited for field encoding. No HVS model parameter adjustments were done because of the limited PQA200 access time we had.

Figs. 6.3, 6.4, 6.5 present the PQA measurement results. The smaller the JND value, the better is its perceptual quality. To understand the scale, it is possible to compare results of Source Encoding to 6Mbps (”Input”) and Source Encoding to 4Mbps (”Encoded”). Re-encoding always provides the lowest quality measured. Encoding of the source sequence is always better than ”Simple” transrating, and can outperform ”Simple with MSE” and even ”Lagrangian” in some cases. In our tests the trellis-based methods always provided better results than the encoding of the original sequence to the same rate.

It is also seen that adding HVS weighting always improves PQA results. Of course,

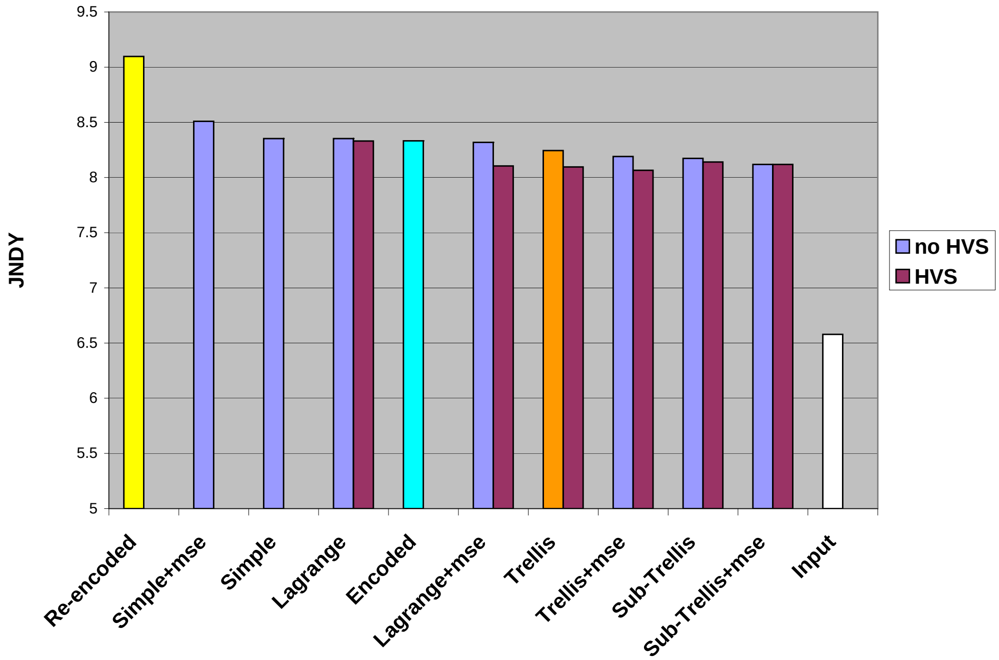


Figure 6.4: Perceptual Evaluation Results: CHEERS sequence transrated from 6Mbps to 4Mbps.

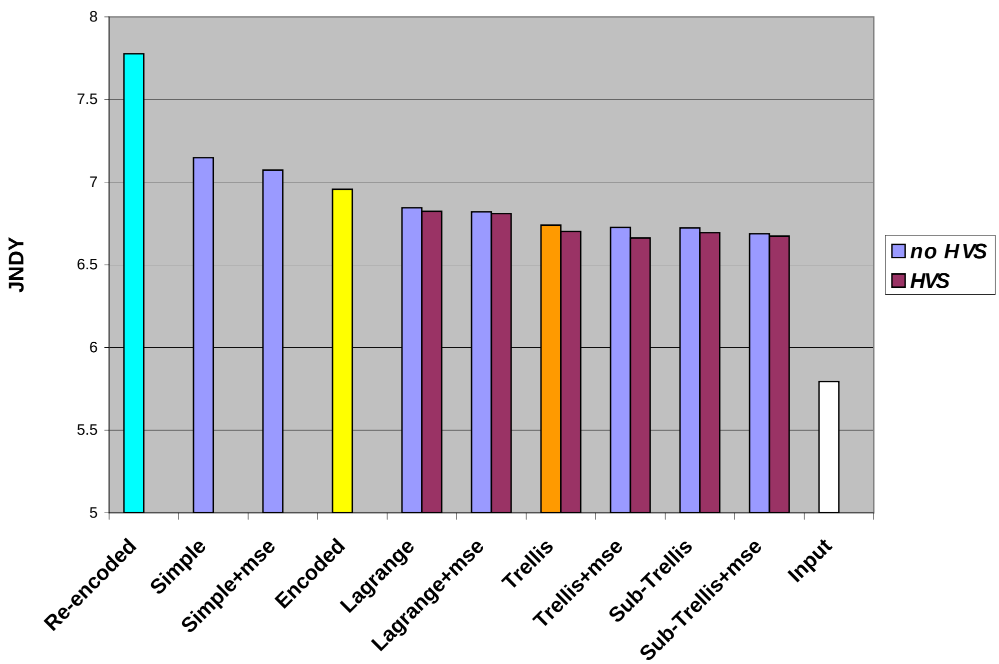


Figure 6.5: Perceptual Evaluation Results: MOBILE sequence transrated from 6Mbps to 4Mbps.

it does not say that true subjective perceptual evaluation results will be the same, but it shows that our assumed HVS model is possibly similar to the model used by the PQA200 tool. We are quite sure that it is possible to improve the results by a better parameter selection, even in the current model, which was not done because of the reasons mentioned above.

Figures 6.6-6.9 show that using perceptual weighting indeed improves the subjective quality of transrated video. As can be seen from the difference pictures presented in Fig.6.8,6.9, the use of perceptual weighting allows the introduction of greater distortion in areas where we can't see it, like the flower field. It thus allows us to provide more bits to areas that affect more our subjective evaluation, like the buildings roofs and the tree trunk, as well as sky details. It is observed that Fig. 6.7 has superior perceptual quality than that of Fig. 6.6, despite of the fact that Fig. 6.6 has a higher PSNR.

Figure 6.6: Subjective Evaluation: GARDEN sequence transrated from 6Mbps to 1.7Mbps without applying HVS weighting (Sub-optimal Trellis method).

Figure 6.7: Subjective Evaluation: GARDEN sequence transrated from 6Mbps to 1.7Mbps with HVS weighting (Sub-optimal Trellis method).

Figure 6.8: Subjective Evaluation: GARDEN sequence transrated from 6Mbps to 1.7Mbps without HVS weighting (Sub-optimal Trellis method), difference from source.

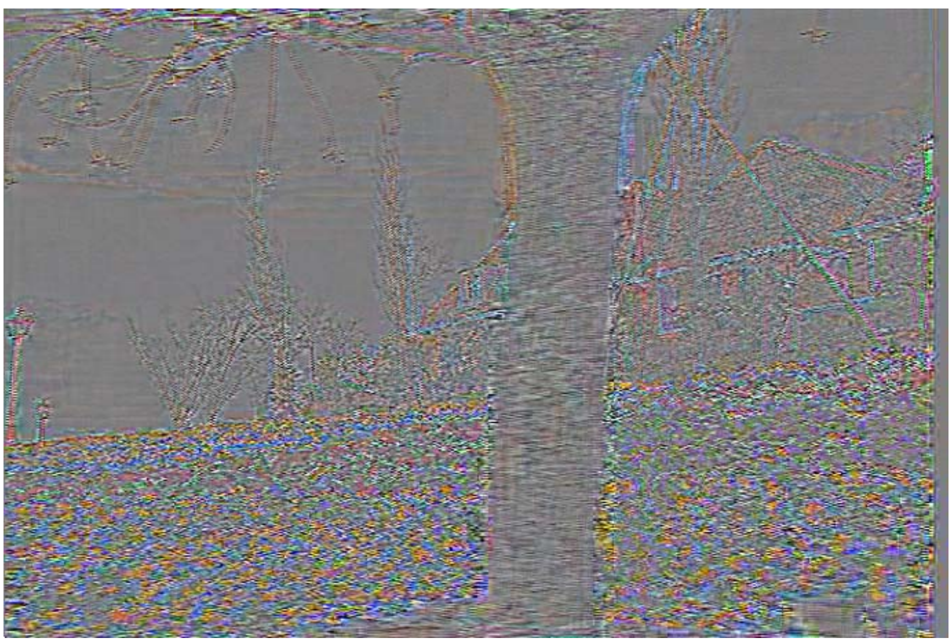


Figure 6.9: Subjective Evaluation: GARDEN sequence transrated from 6Mbps to 1.7Mbps with HVS weighting (Sub-optimal Trellis method), difference from source.

Chapter 7

Conclusions and Future Directions

7.1 Summary and Conclusions

With the wide spread of multimedia services that provide video over heterogeneous networks, transrating becomes an important part of transmission. Transrating allows matching pre-encoded video parameters with bandwidth constraints of each end-user connection.

In this thesis different issues related to transrating have been investigated. The goal of the current research is to present the state of the art transrating techniques, and to develop a high quality video transrating system. In order to clarify the contributed work, we provide here the block diagram of the final scheme and will summarize our contribution in its relevant parts.

On Fig. 7.1 the block diagram of the proposed transrating system is presented. The following notations are used:

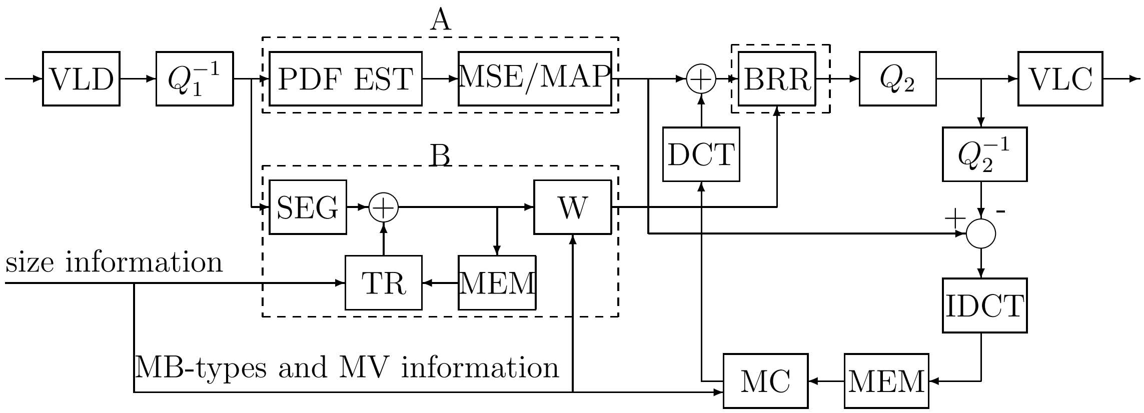


Figure 7.1: Proposed Transrating scheme

VLD	-	Variable Length Decoding
VLC	-	Variable Length Encoding
$Q_{1/2}^{-1}$	-	Inverse Initial/Secondary Quantization
Q_2	-	Secondary Quantization
PDF EST	-	DCT coefficients PDF estimation block
MSE/MAP	-	MAP/MSE requantization error reduction algorithms
BRR	-	Bit-Rate Reduction via requantization
IDCT/DCT	-	Inverse/direct Digital Cosine Transform
MEM	-	Memory buffer for Tracking and Error Compensation
MC	-	Motion Compensation
TR	-	Tracking based on MV information
SEG	-	DCT-domain Segmentation Algorithm
W	-	Class and Motion Weighting block
A	-	Statistical-based quality improvement unit
B	-	Segmentation and Tracking unit

The proposed system is an extension of Fast Pixel Domain transcoder (see Fig.2.3). Dashed boxes shows the functional blocks we contributed to.

The first and the most important part of this work is the development and evaluation of different requantization algorithms. This thesis focuses on frame-level requantization. In chapter 2 a fast complexity-based transrating algorithm is presented. For this and all the following algorithms any GOP-level rate controlling scheme can be applied.

Chapter 4 introduces Lagrangian optimization for quantization step-size selection developed in [5]. This algorithm has provided the best PSNR for given bit-rate constraints over all previously developed transrating algorithms.

We propose in this work to extend the Lagrangian optimization procedure by allowing the modification of quantized DCT indices, including setting their values to zero, in addition to quantization step-size selection. Quantized DCT index-modification and quantization step-size selection are optimally done using a low complexity trellis-based procedure. The proposed requantization algorithm provides higher PSNR values than the original Lagrangian-based optimization method that only handles the selection of quantization steps, and still, practically, does not exceed considerably its complexity.

The main problem of transrating is that the input video is already degraded by lossy compression. In chapter 3 DCT coefficient PDF-based methods for requantization error minimization are discussed (see dashed box A, Fig.7.1), and the MSE and MAP estimation methods are compared. It is commonly assumed that Laplacian distribution is a good model for DCT coefficients distribution, and a number of methods to estimate Laplacian distribution parameter were proposed in the literature. However, we found that those methods do not provide good results, and developed another method for PDF estimation, from quantized data, by linear interpolation. Our method has another advantage over Laplacian parameter estimation. In our method, all the input coded coefficients are utilized, while Laplacian estimators can work only with one quantization step-size in the input data set. Using the proposed method results, on average, in superior performance to Laplacian distribution-based estimation.

To take into account HVS properties, chapter 6 provides an extension to the Lagrangian-based methods (see dashed box B, Fig.7.1). We developed a segmentation algorithm that uses quantized DCT coefficients values to classify each block as either smooth, textured or boundary. Because the difference pictures do not have sufficient information for segmentation, we implemented a block tracking system to utilize the segmentation information from previously processed frames. Temporal masking is used by applying additional weighting to fast moving blocks.

The proposed system is an efficient transrating scheme that provides the best documented results todate, in term of PSNR, and demonstrates a way to further improve it by taking into consideration HVS properties.

7.2 Future Directions

This thesis shows that it is possible to achieve improved performance (in terms of PSNR) by changing quantized coefficient indices, and not only the quantization step-sizes, at a tolerable complexity cost. In this work we focused on frame-level rate reduction schemes, mainly for MPEG-2 transrating. However, the proposed trellis-based algorithm can be applied to other coding standards that imply run-level VLC. It is an open issue for further examination if it can be used for the H.264 standard uses a different coefficients encoding scheme.

The proposed optimization can be applied not only for transcoding, but also to encoding of the initial video sequence, to improve the encoding results.

We also did not address the optimization of GOP-level transcoding schemes, which is an important issue for building transrating systems.

There are many other rate reduction techniques such as spatial resolution reduction, frame dropping and others that can be combined with the solutions proposed in the current work, and need to be studied.

We propose in this work a simple approach for taking in consideration HVS characteristics in existing PSNR-based transrating algorithms. However, parameter values (weights) need to be determined on the basis of subjective studies, as well as more complex HVS models need to be utilized in order to further improve perceptual quality of transrating systems, as well as of encoding algorithms.

Appendix A

Results for MAP/MSE

Additional simulation results for chapter 3 are presented here.

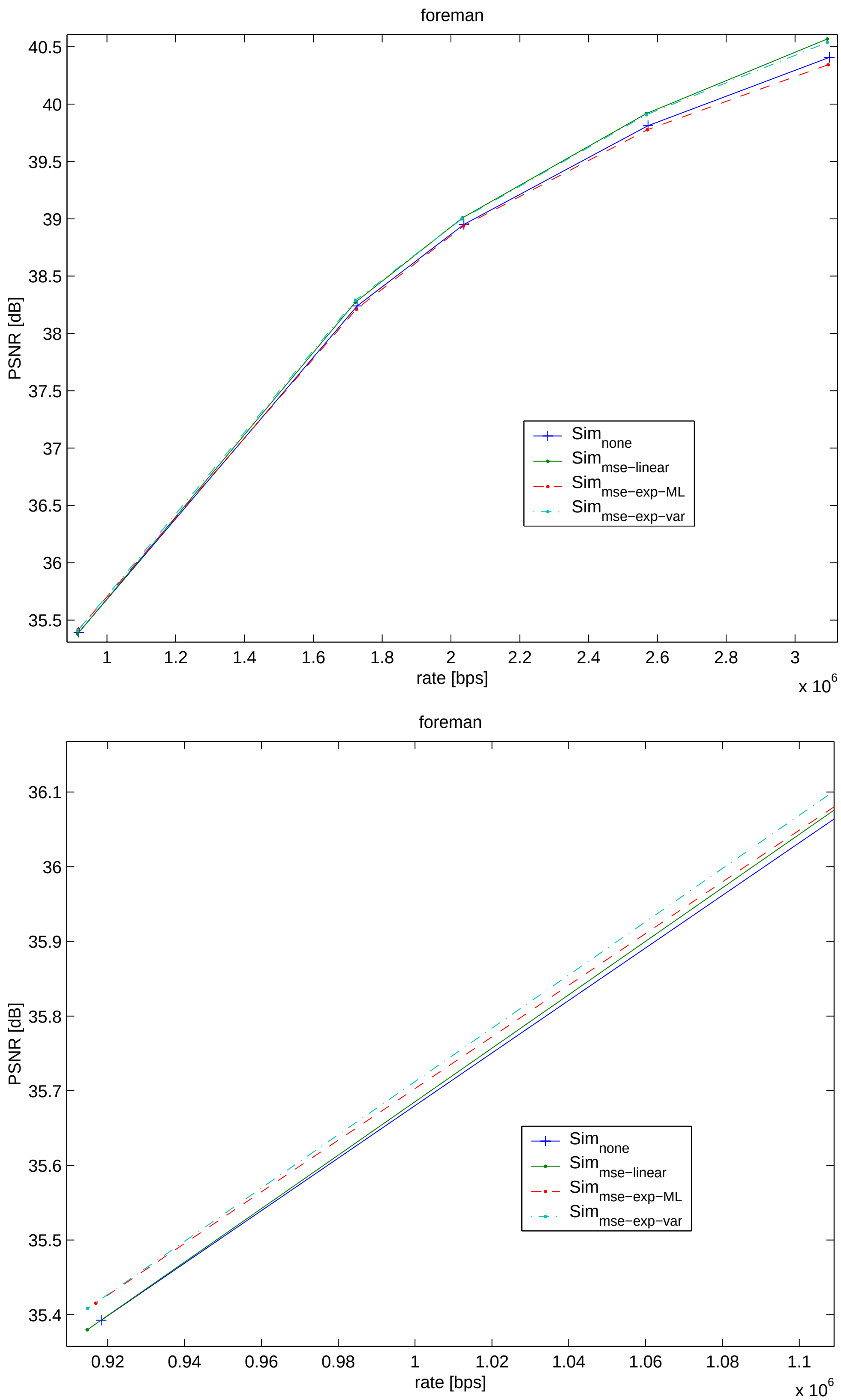


Figure A.1: MSE simple transcoding, FOREMAN sequence.

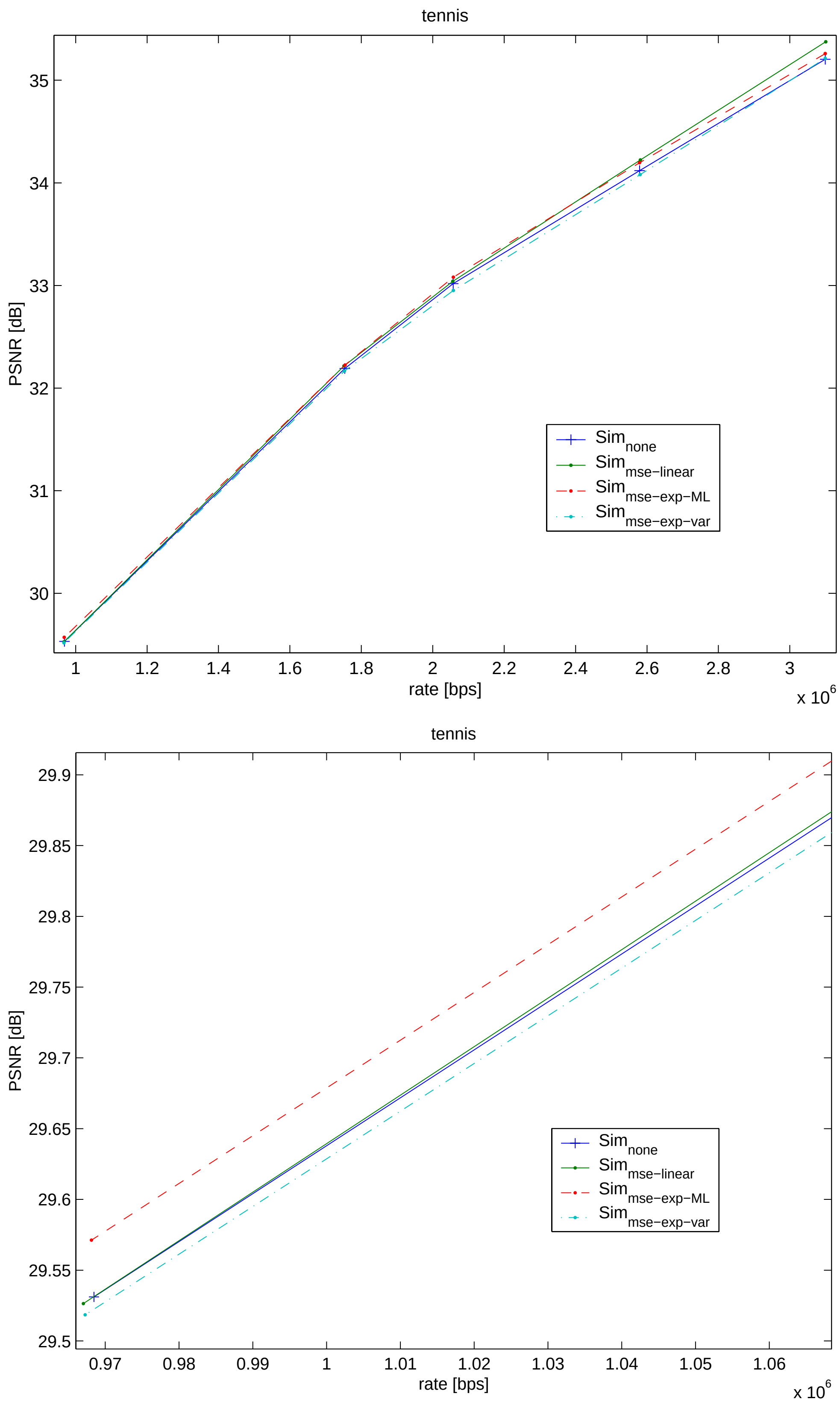


Figure A.2: MSE simple transcoding, TENNIS sequence.

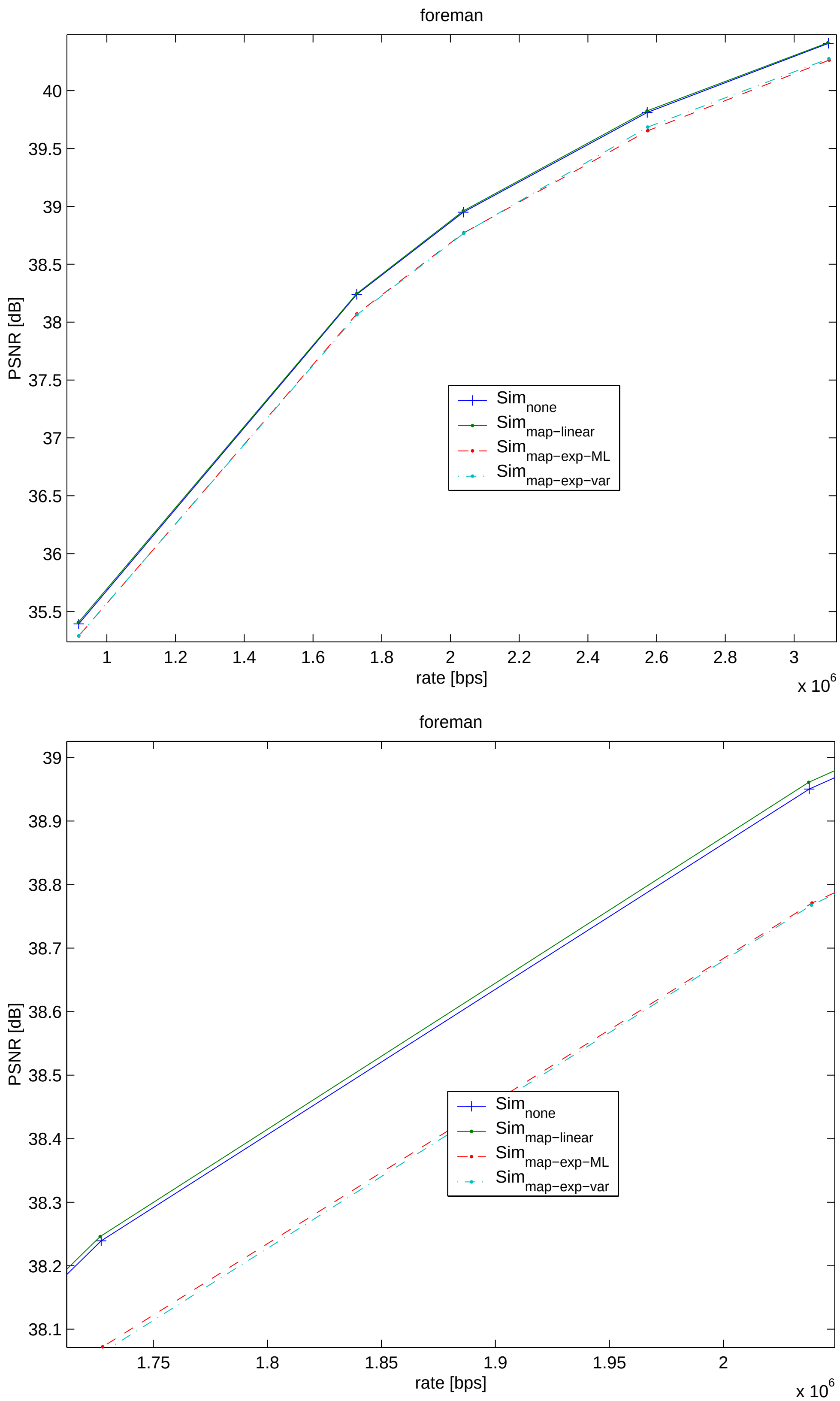
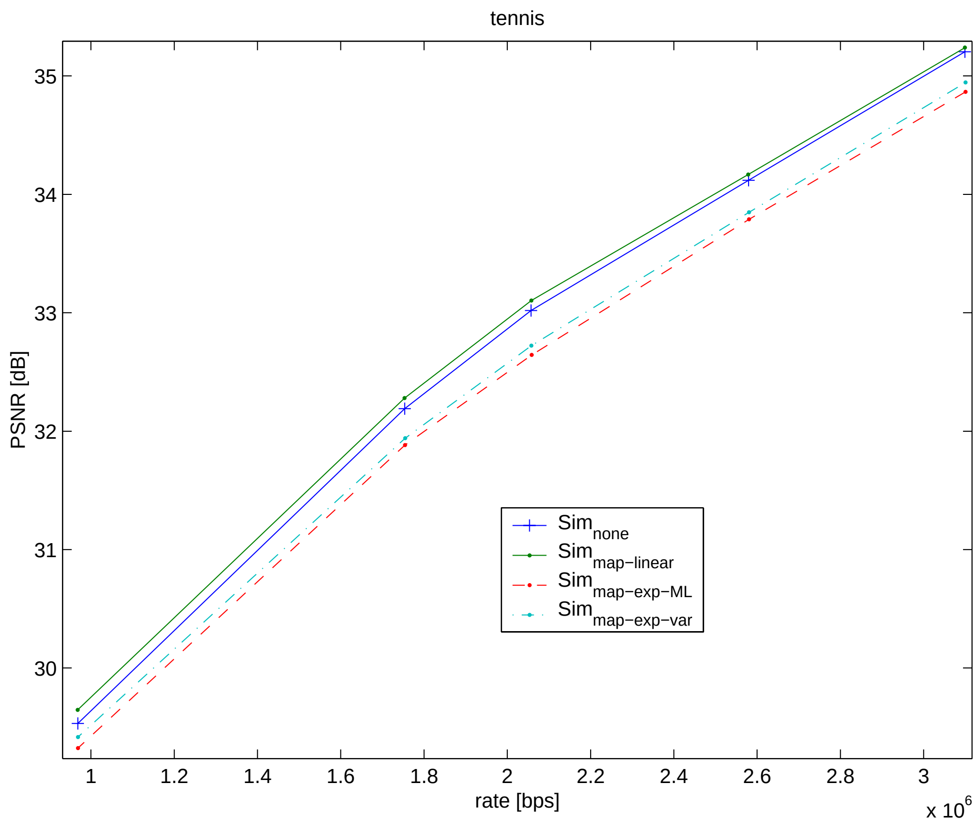


Figure A.3: MAP simple transcoding, FOREMAN sequence.



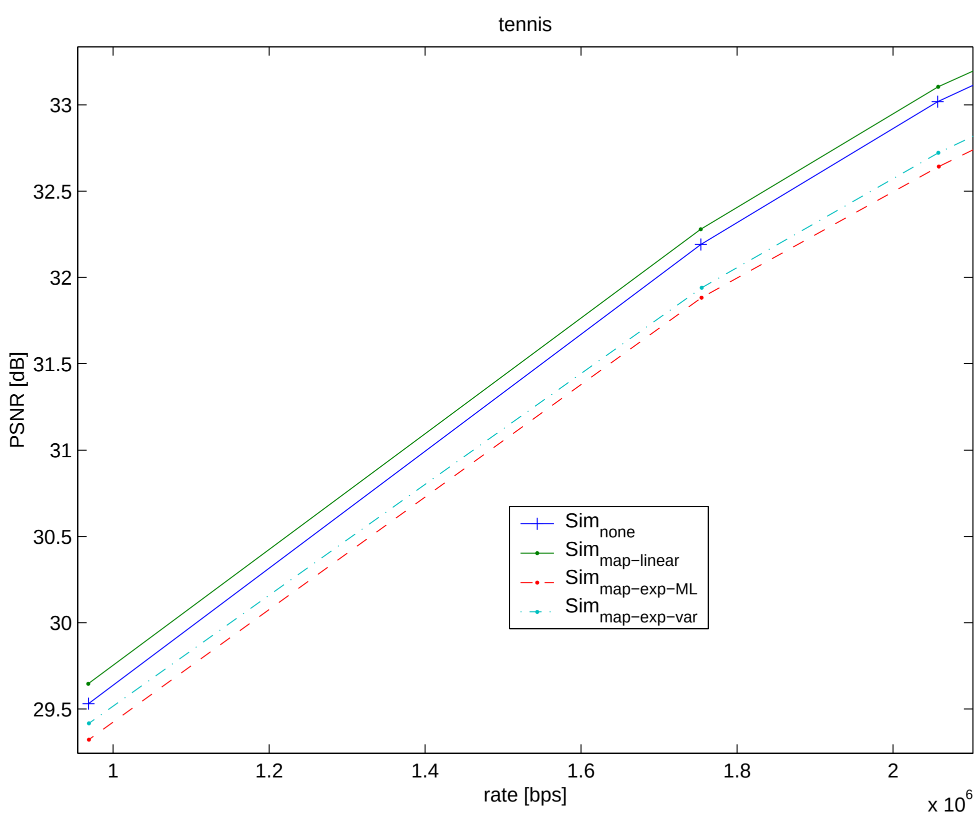


Figure A.4: MAP simple transcoding, TENNIS sequence.

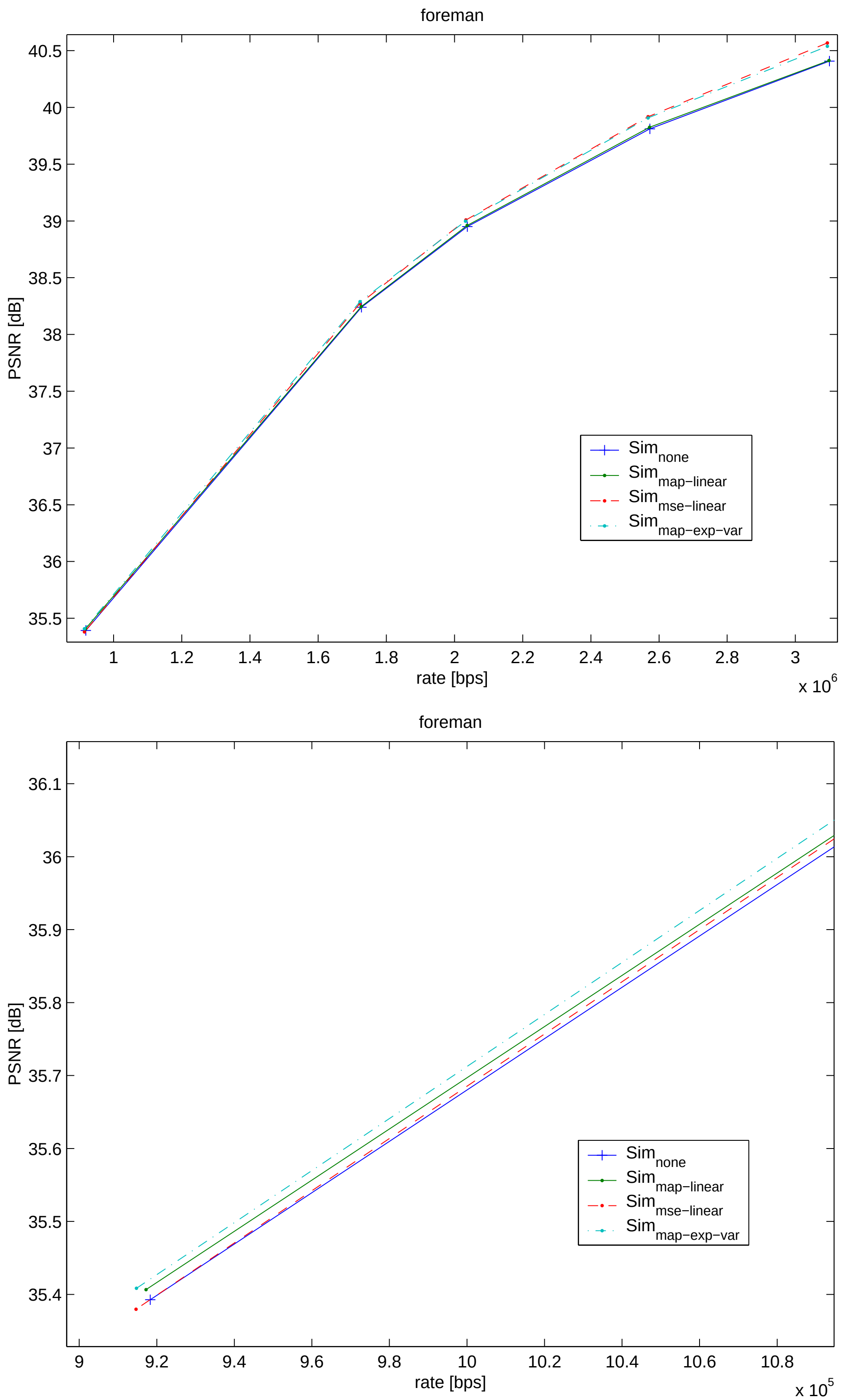
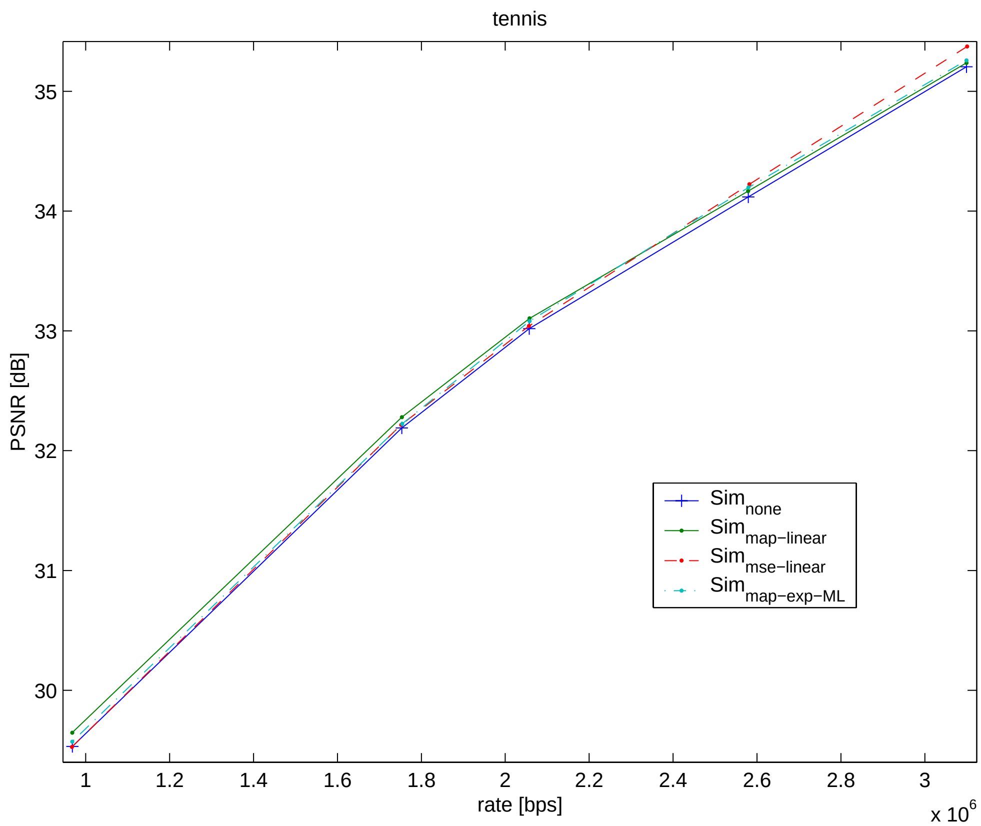


Figure A.5: compare MSE and MAP for simple transcoding, FOREMAN sequence.



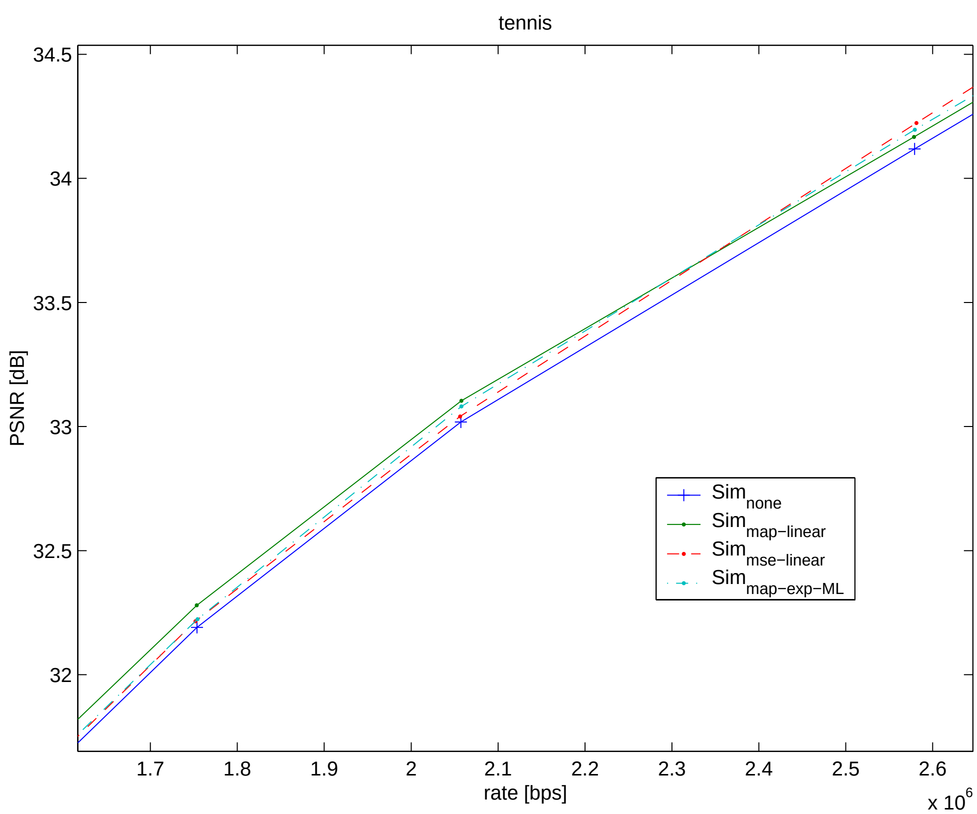


Figure A.6: compare MSE and MAP for simple transcoding, TENNIS sequence.

Appendix B

Results for Lagrangian optimization

Additional simulation results for chapter 4 are presented here.

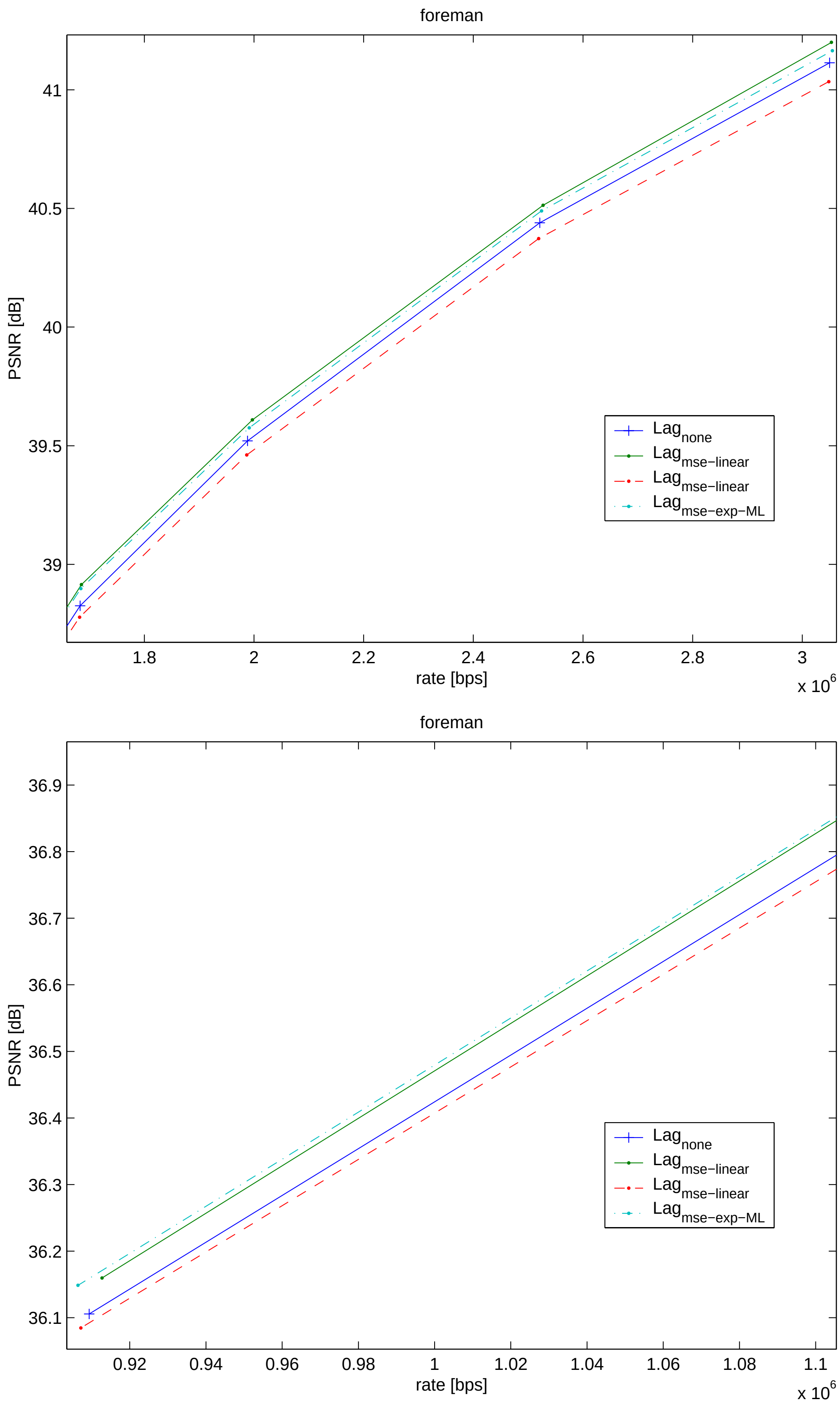
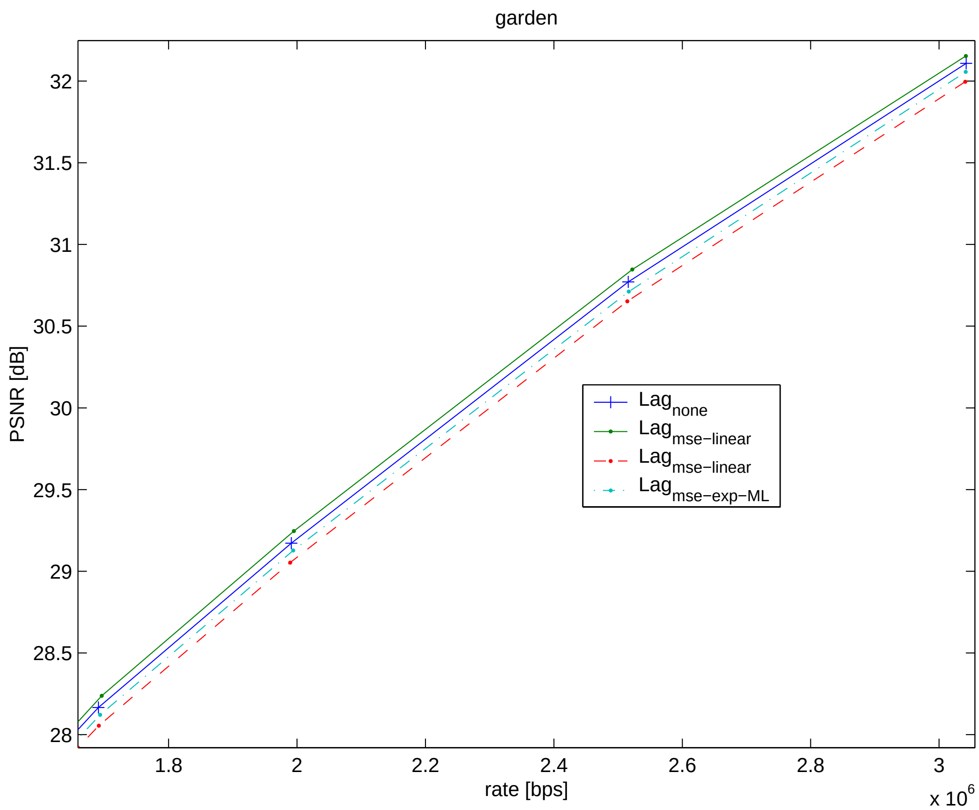


Figure B.1: MSE Lagrangian transcoding, FOREMAN sequence.



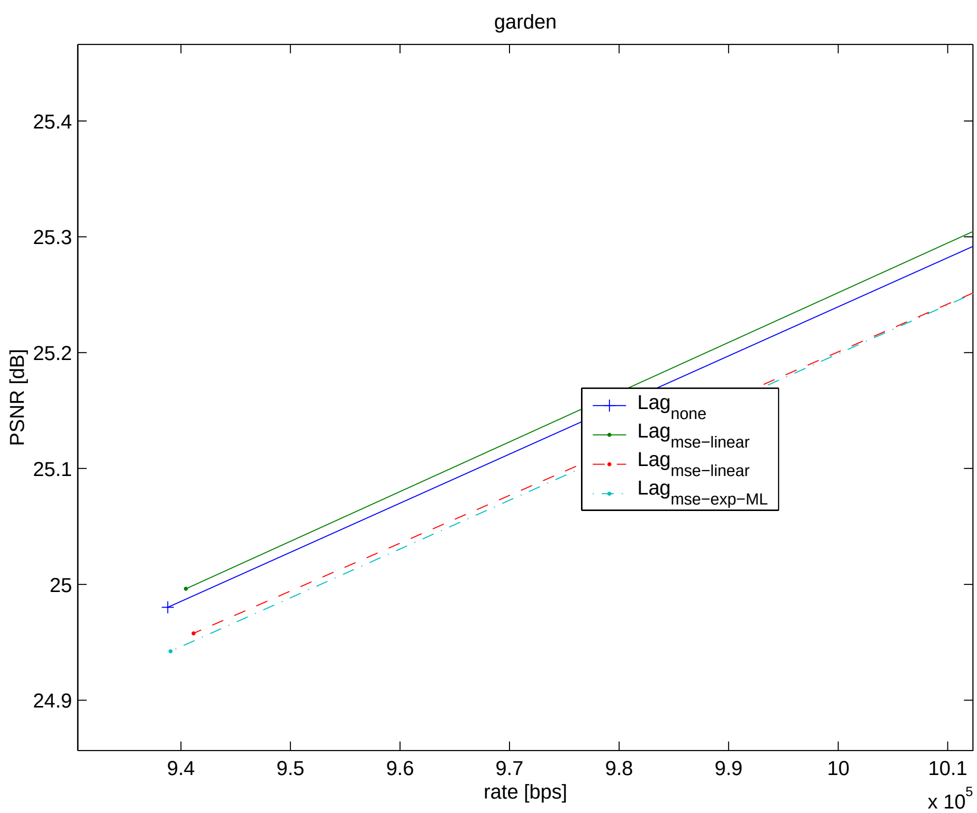
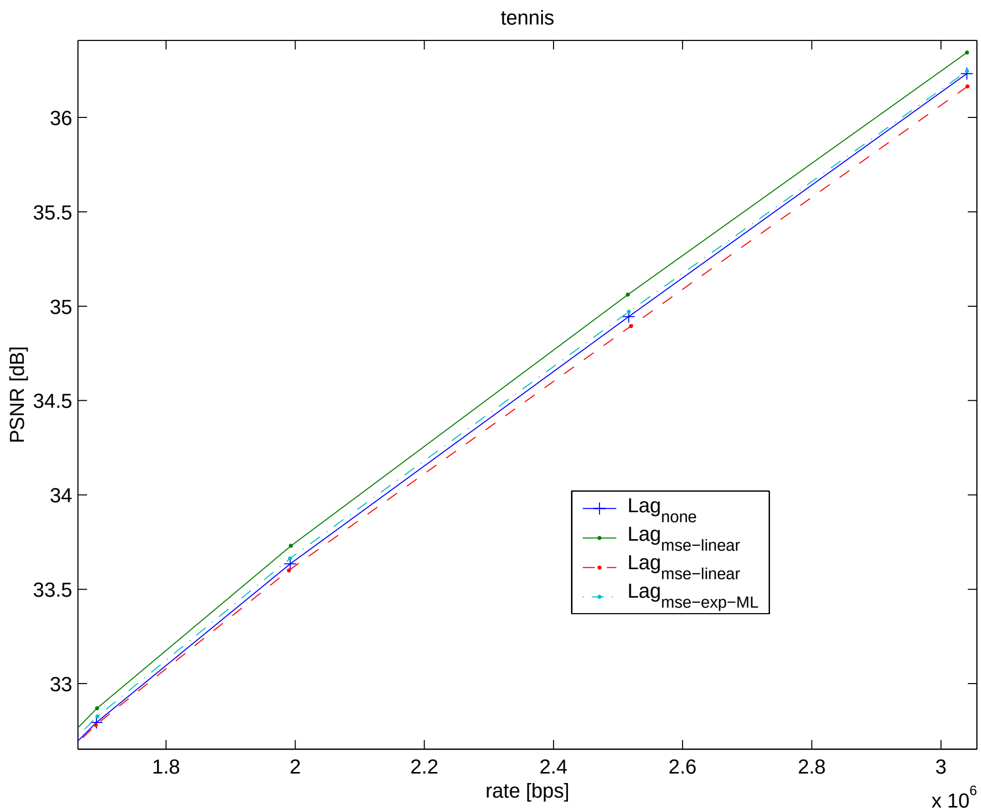


Figure B.2: MSE Lagrangian transcoding, GARDEN sequence.



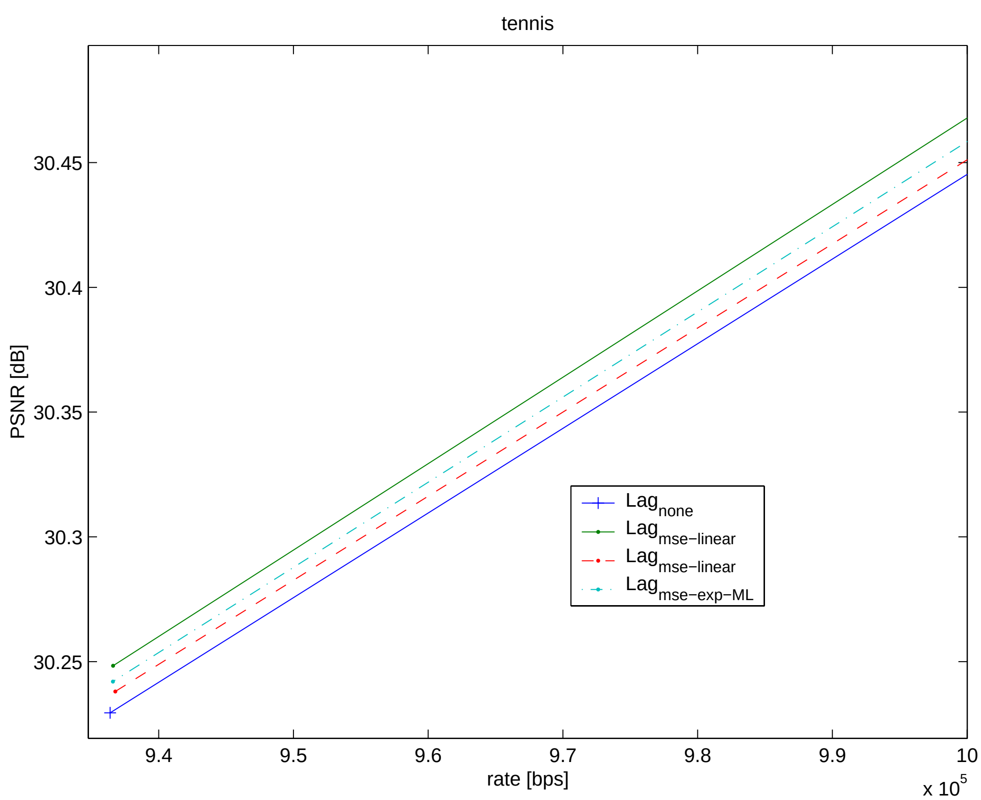


Figure B.3: MSE Lagrangian transcoding, TENNIS sequence.

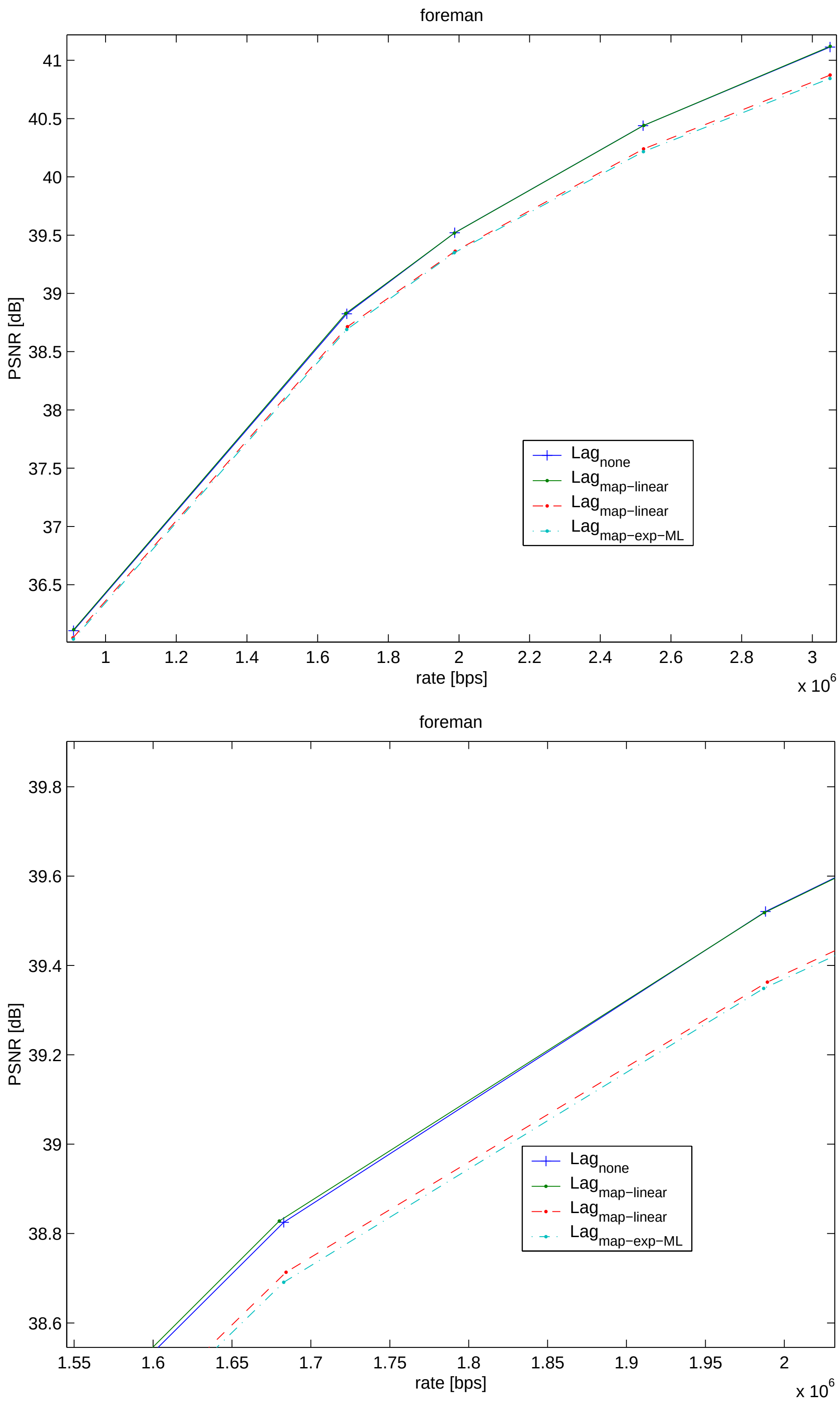
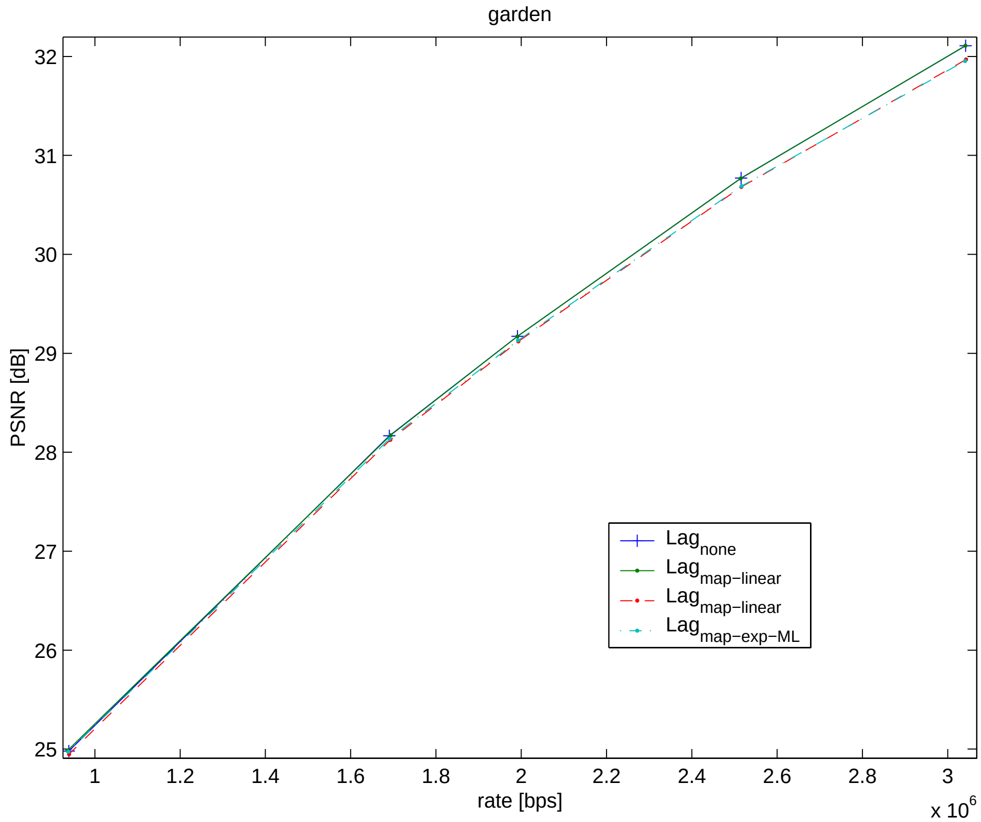


Figure B.4: MAP Lagrangian transcoding, FOREMAN sequence.



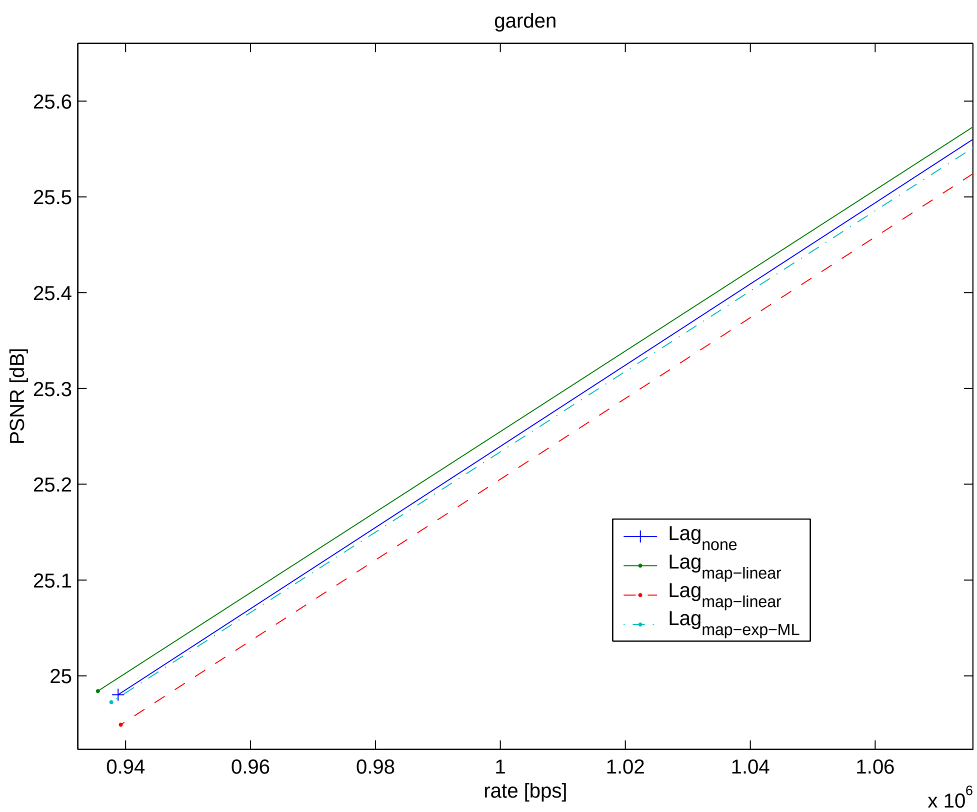


Figure B.5: MAP Lagrangian transcoding, GARDEN sequence.

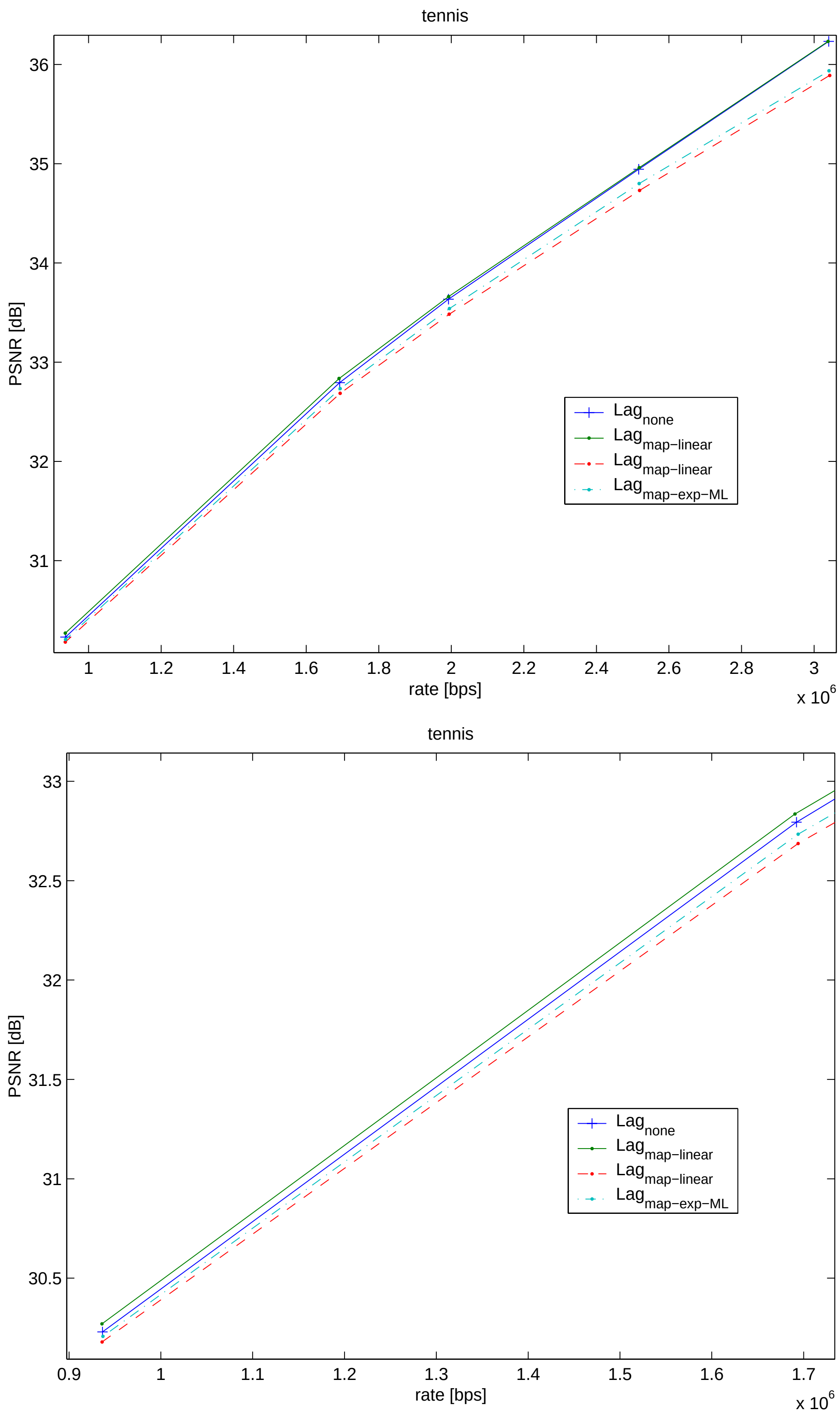
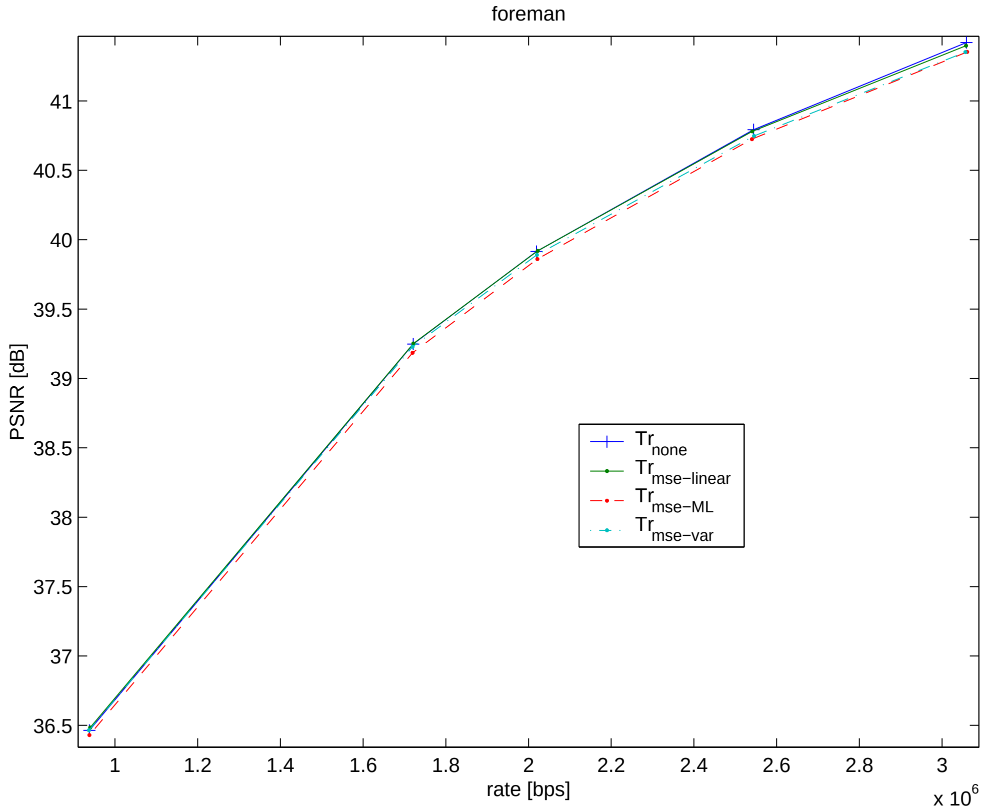


Figure B.6: MAP Lagrangian transcoding, TENNIS sequence.

Appendix C

Results for Extended Lagrangian optimization

Additional simulation results for chapter 5 are presented here.



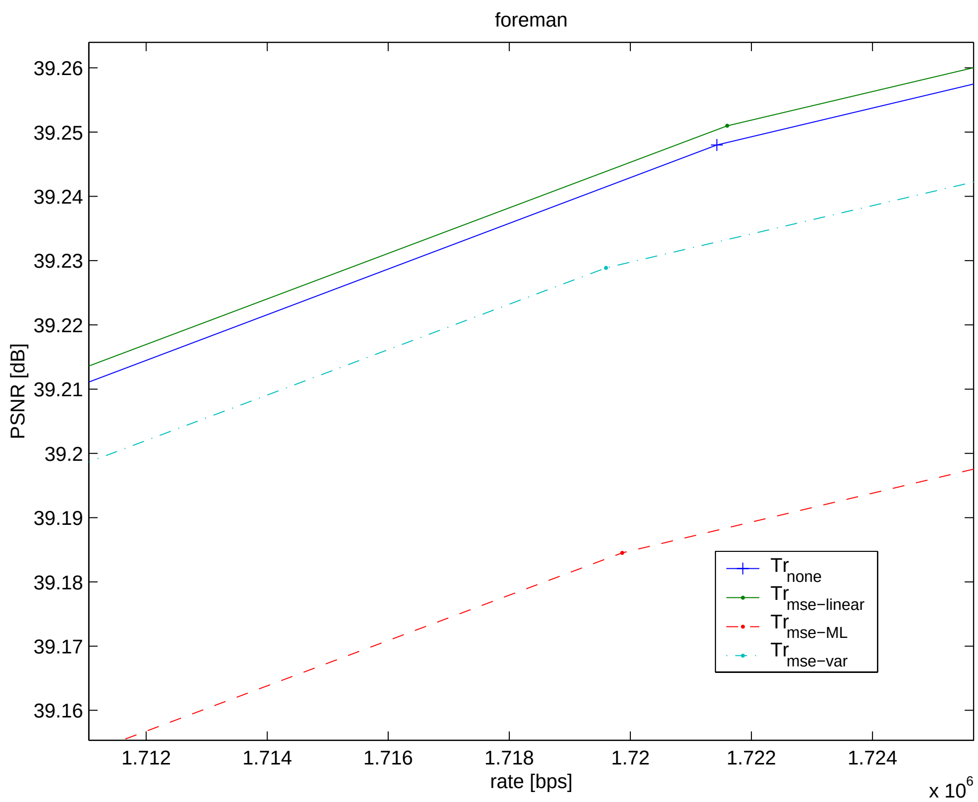
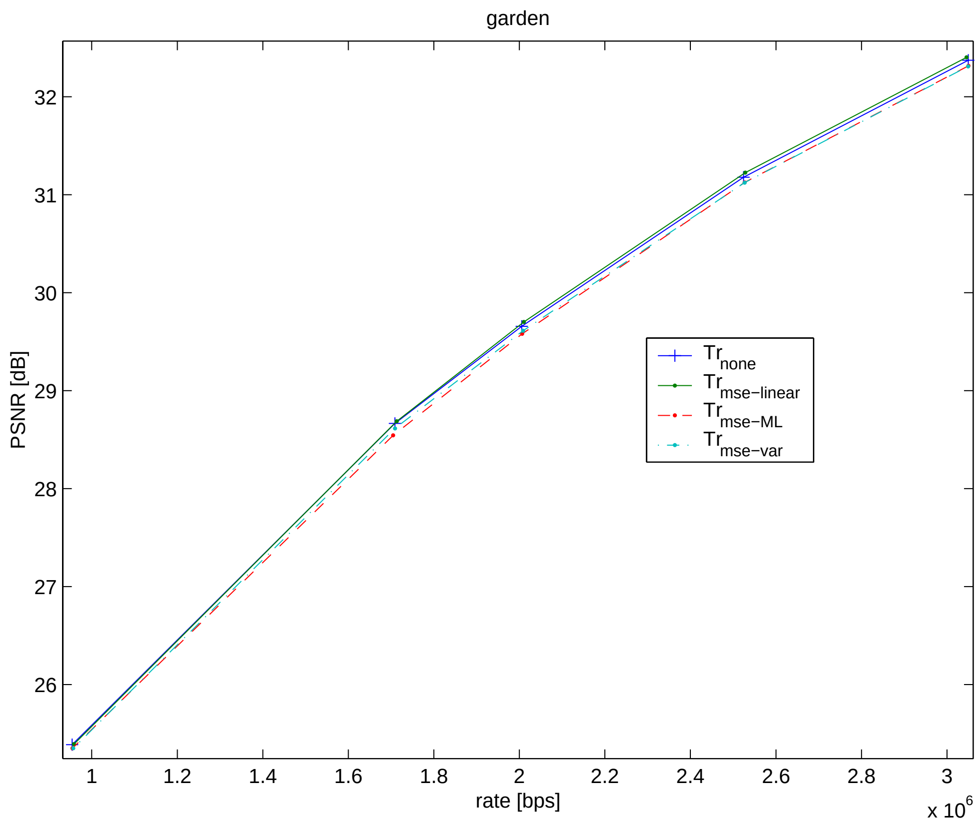


Figure C.1: Extended Lagrangian transcoding, FOREMAN sequence.



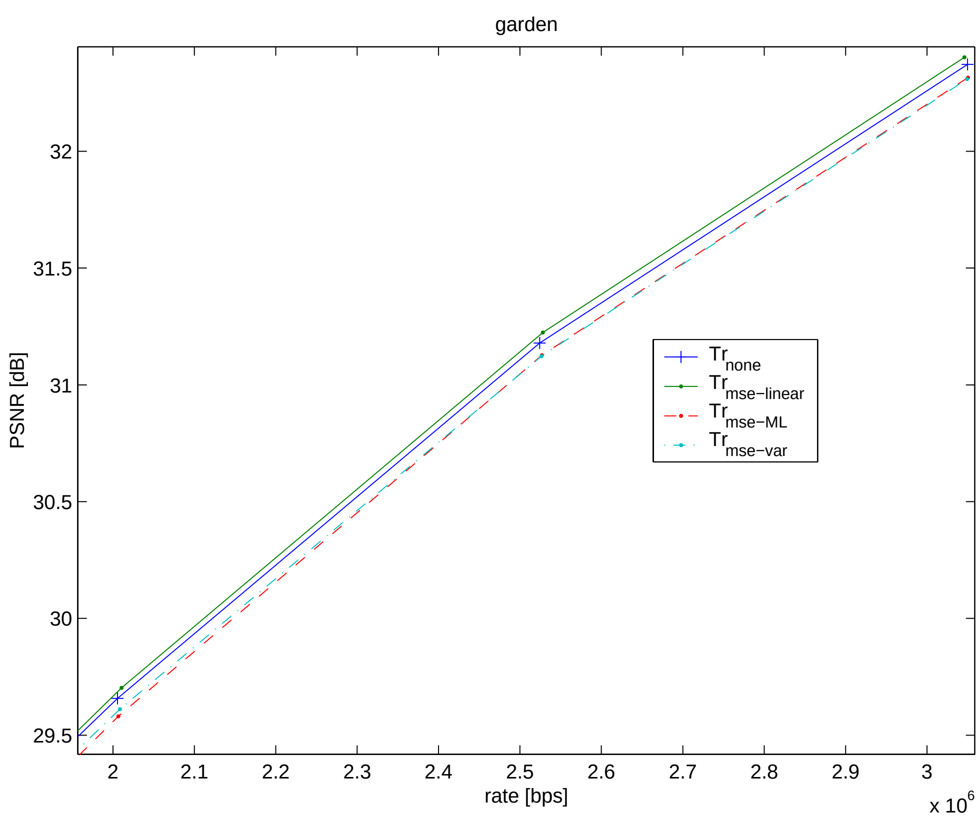
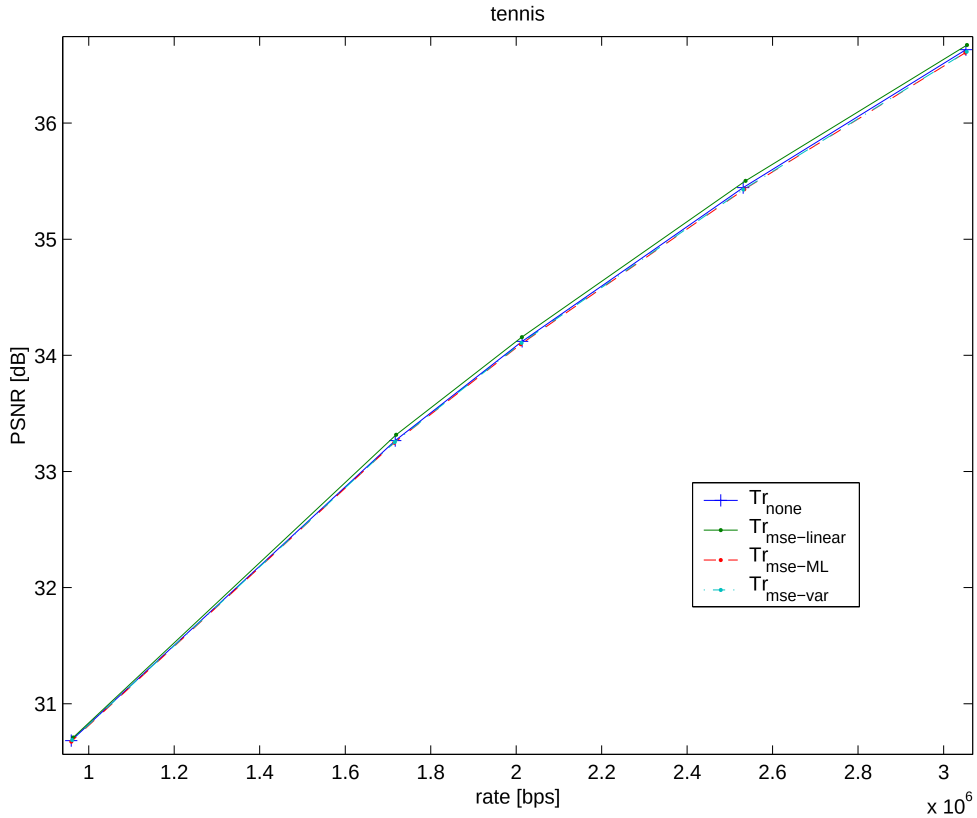


Figure C.2: Extended Lagrangian transcoding, GARDEN sequence.



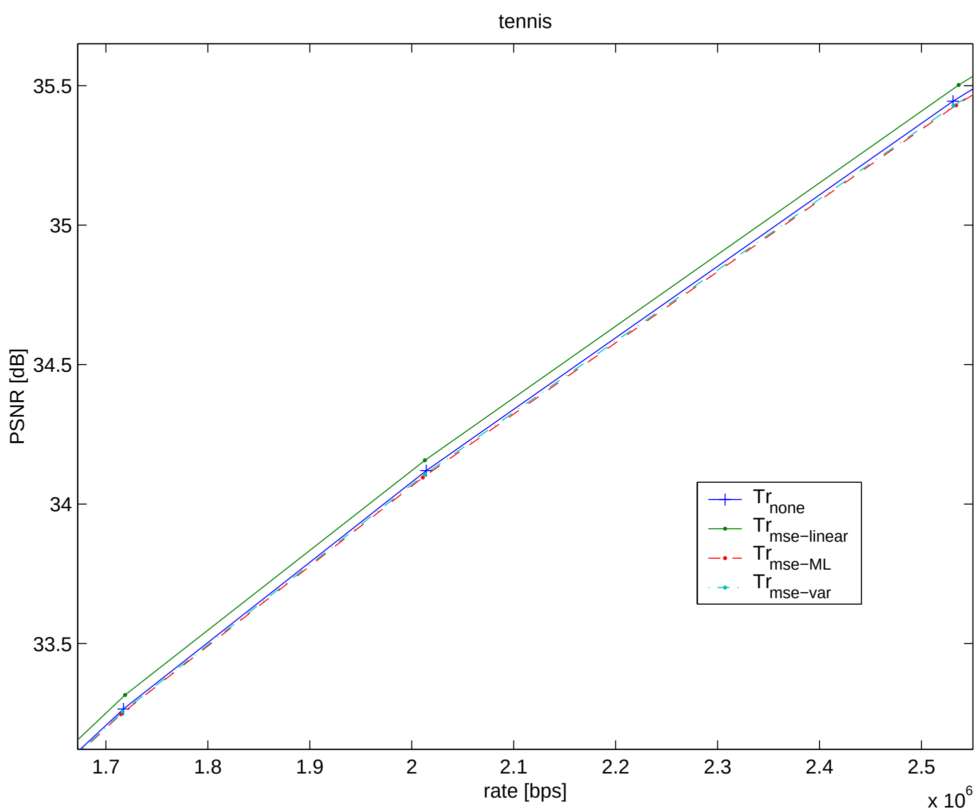
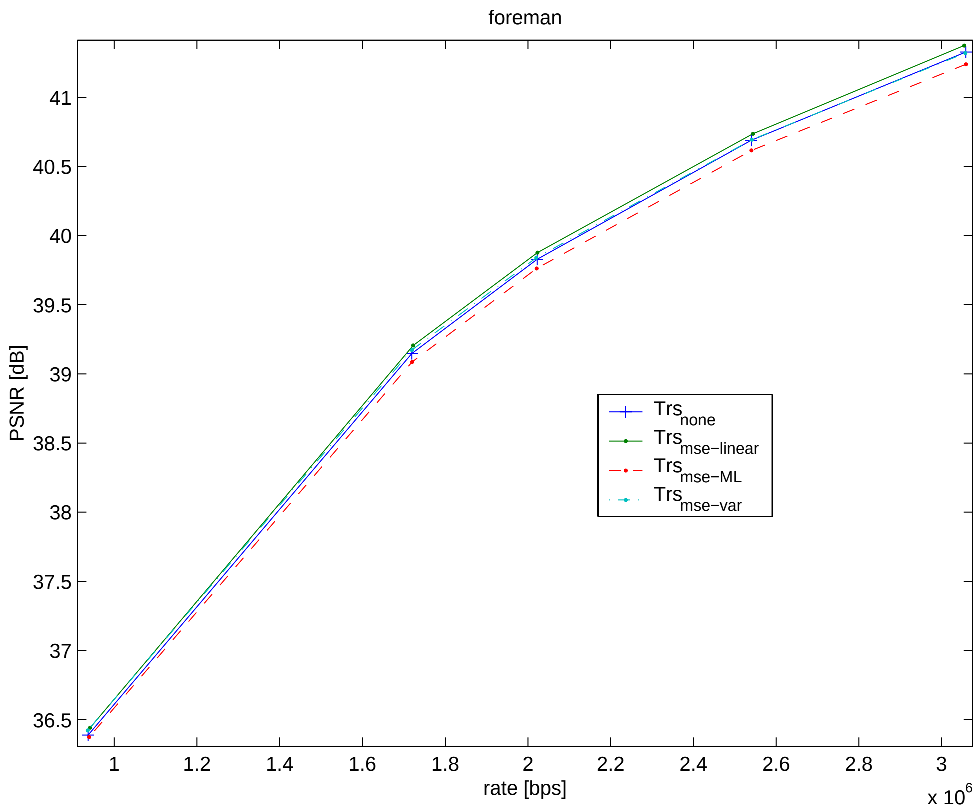


Figure C.3: Extended Lagrangian transcoding, TENNIS sequence.



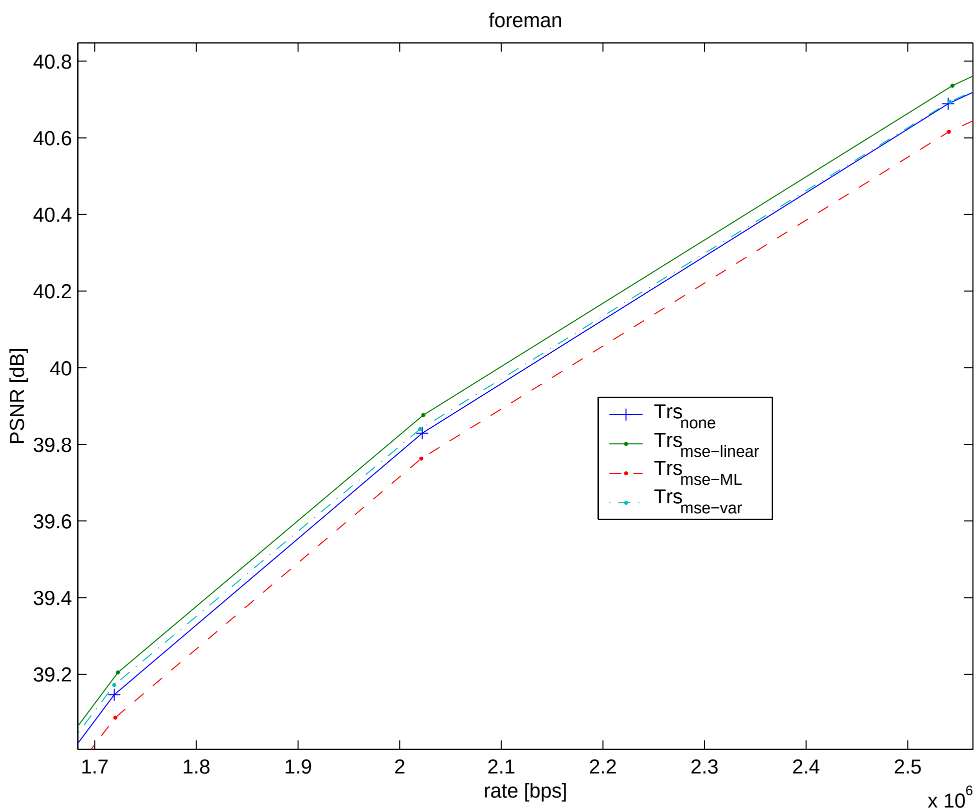


Figure C.4: Sub-optimal Extended Lagrangian transcoding, FOREMAN sequence.

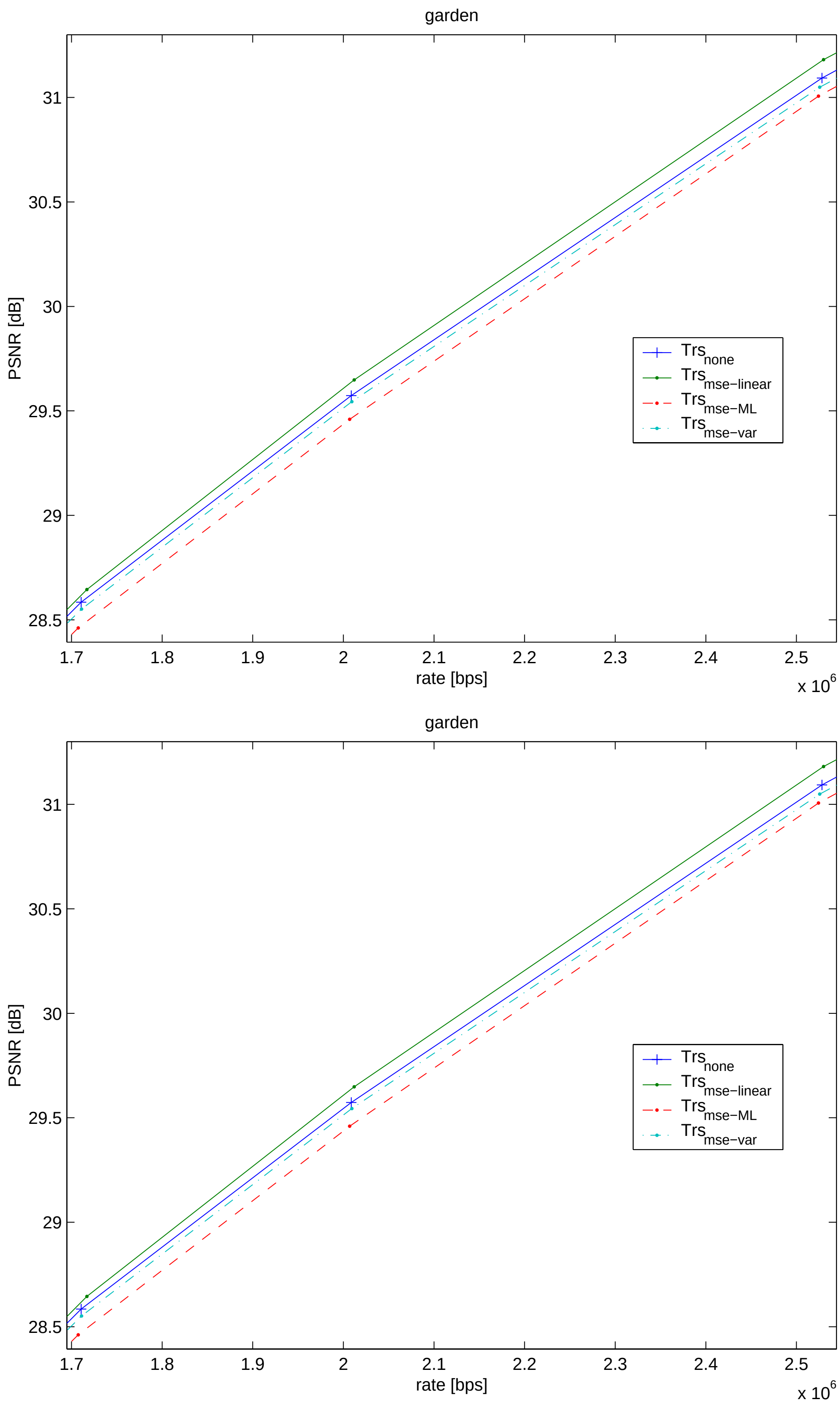
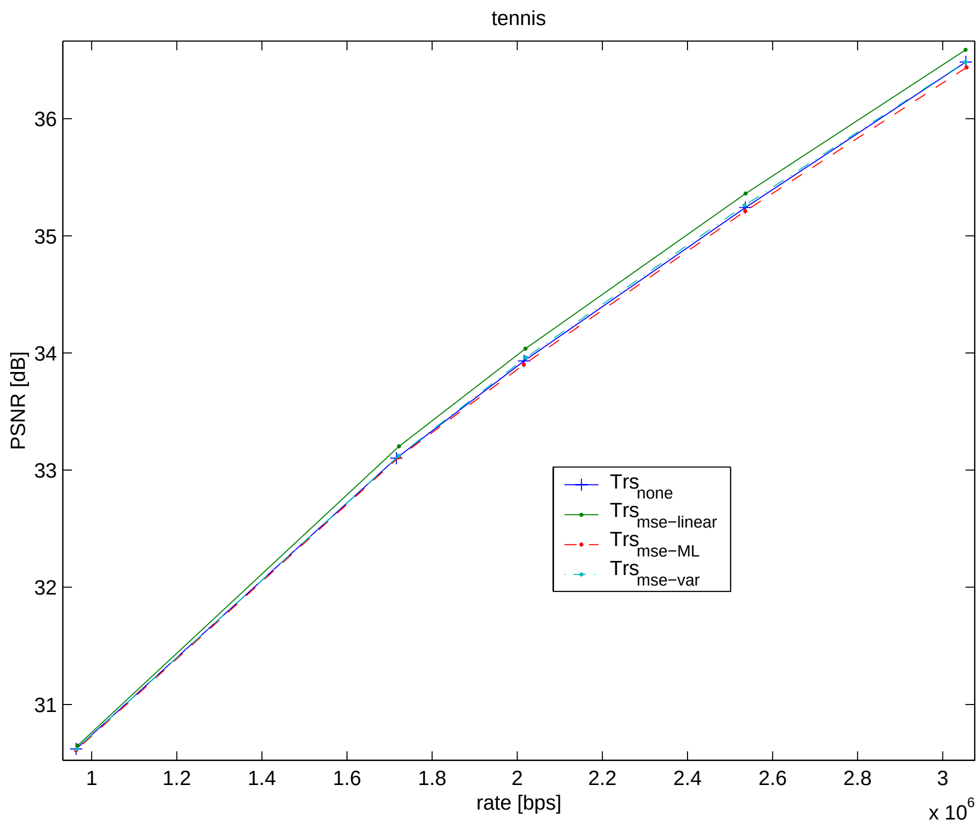


Figure C.5: Sub-optimal Extended Lagrangian transcoding, GARDEN sequence.



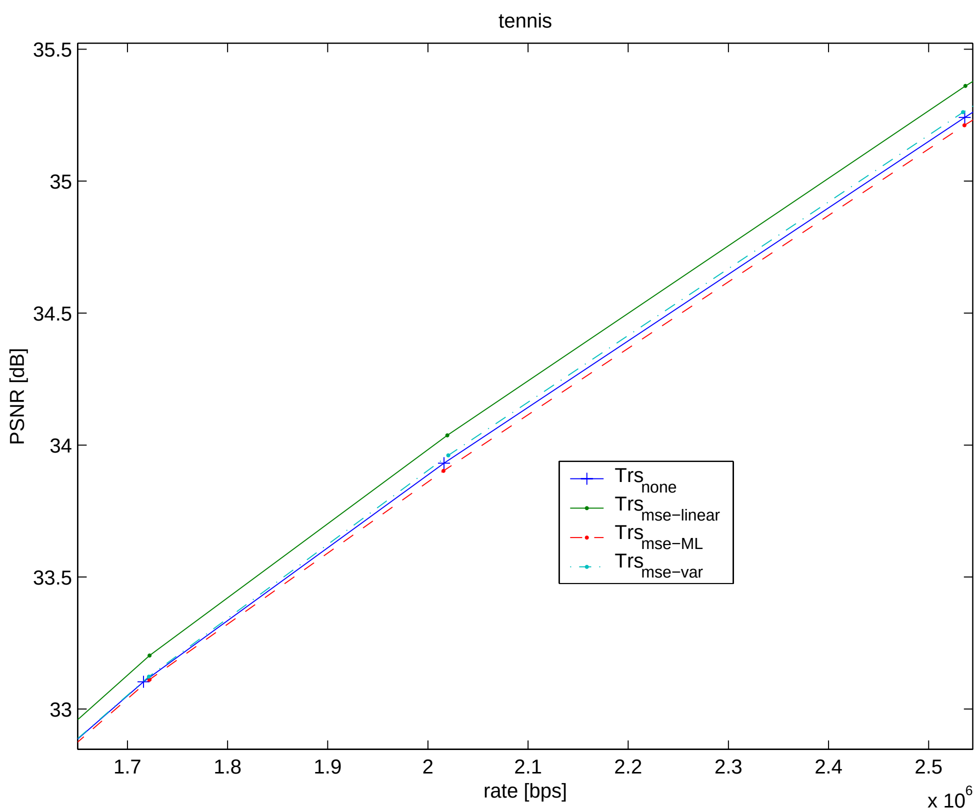


Figure C.6: Sub-optimal Extended Lagrangian transcoding, TENNIS sequence.

Bibliography

[1] H. Sorial, W. E. Lynch, A. Vincent, "Joint transcoding of multiple MPEG video bitstreams," *Circuits and Systems, 1999. ISCAS '99. Proceedings of the 1999 IEEE International Symposium on*, vol. 4, pp. 251 –254, 1999.

[2] J.-N. Hwang, T.-D. Wu, C.-W. Lin , "Dynamic frame-skipping in video trasnscoding," *Multimedia Signal Processing, 1998 IEEE Second Workshop on*, pp. 616 –621, 1998.

[3] G. Keesman et all , "Transcoding of MPEG bitstreams," *Signal Processing: Image Communication*, vol. 8, pp. 481–500, 1996.

[4] A. A. Assunção, M. Ghanbari , "A frequency-domain video transcoder for dynamic bit-rate reduction of MPEG-2 bit stream," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 8, pp. 953 –967, Dec.1998.

[5] A. A. Assunção, M. Ghanbari , "Optimal bit rate conversion of MPEG-2 video bit streams," *Electronics Letters*, vol. 33, pp. 675–677, Apr.1997.

[6] B. Tao, H. A. Peterson, B. W. Dickinson, "A rate-quantization model for MPEG encoders," *Image Processing, 1997. Proceedings., International Conference on*, vol. 1, pp. 338 –341, 1997.

[7] Y. Nakajima, H. Hori, T. Kanoh, “Rate conversion of MPEG coded video by re-quantization process,” *Image Processing, 1995. Proceedings., International Conference on*, vol. 3, pp. 408 –411, 1995.

[8] H. Sun, W. Kwok, J. W. Zdepski, “Architecture for MPEG compressed bitstream scaling,” *IEEE trans. on circuits and systems for video technology*, vol. 6, no. 2, pp. 191–199, Apr. 1996.

[9] A. A. Assunção, M. Ghanbari, “Transcoding of single-layer MPEG video into lower rates,” *Vision, Image and Signal Processing, IEE Proceedings*, vol. 144, pp. 377 –383, Dec. 1997.

[10] A. A. Assunção, M. Ghanbari, “Transcoding of MPEG-2 video in the frequency domain,” *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol. 4, pp. 2633 –2636, Dec. 1997.

[11] S. Acharya, B. Smith, “Compressed domain transcoding of MPEG,” *Multimedia Computing and Systems, 1998. Proceedings. IEEE International Conference on*, p. 295 304, 1998.

[12] N. Björk, C. Christopoulos, “Transcoder architectures for video coding,” *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 5, pp. 2813–2816, 1998.

[13] N. Björk, C. Christopoulos, “Transcoder architectures for video coding,” *IEEE Transactions on Consumer Electronics*, vol. 44, no. 1, pp. 88–98, February 1998.

[14] K. Lee, H. S. Chang, S. S. Chun, H. Choi, S. Sull, “Perception-based image transcoding for universal multimedia access,” *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 2, pp. 475 –478, 2001.

[15] M. Plotnick, T. Kinoshita, H. Yoshigi, "High-quality bit-rate transformation for the home media station," *Consumer Electronics, 1998. ICCE. 1998 Digest of Technical Papers. International Conference on*, pp. 102 –103, 1998.

[16] R. C. Reininger, J. D. Gibson , "Distribution of the two-dimensional DCT coefficients for images," *IEEE trans. Communication*, vol. COM-31, no. 6, pp. 835–839, June 1983.

[17] A. Eleftheriadis, D. Anastassioiu , "Constrained and general dynamic rate shaping of compressed digital video," *Image Processing, 1995. Proceedings., International Conference on*, vol. 3, pp. 396 –399, 1995.

[18] T. Shanableh, M. Ghanbari, "Transcoding architectures for DCT-domain heterogeneous video transcoding," *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 1, p. 433436, 2001.

[19] K. -S. Kan, K. -C. Fan, Y. -H. Huang , "Low-complexity and low-delay video transcoding for compressed MPEG-2 bitstream," *Consumer Electronics, 1997. ISCE '97., Proceedings of 1997 IEEE International Symposium on*, pp. 99 –102, 1997.

[20] C. A. Gonzales, E. Viscito , "Motion video adaptive quantization in the transform domain," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 1, pp. 374 –378, Dec. 1991.

[21] J. Xin, M.-T. Sun, K. Chun, "Bit-allocation for transcoding of pre-encoded video streams," *Proc. SPIE, Visual Communications and Image Processing*, vol. 4671, pp. 164–171, 2002.

[22] W.-N. Lie, Y.-H. Chen , "Dynamic rate control for MPEG-2 bit stream transcoding," *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 1, pp. 477–480, 2001.

[23] A. Leventer, M. Porat, "On bit allocation in video coding and transcoding," *IEE International Conference on Visual Information Engineering - VIE 2003*, 2003.

[24] L. Wang, A. Luthra, B. Eifrig, "Adaptive rate control for mpeg transcoder," *ICIP 99. Proc.*, vol. 4, pp. 266 –270, 1999.

[25] L. Wang, A. Luthra, B. Eifrig, "Rate control for mpeg transcoders," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, pp. 222–234, Feb 2001.

[26] K.N. Ngan, T. Meier, C. Zhenzhong, "Improved single-video-object rate control for mpeg-4," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, pp. 385–393, May 2003.

[27] K. Wang and J. W. Woods, "MPEG2 transcoding," *CDVMR PR 98-01*, 1998.

[28] S. Kadono, M. Etoh, N. Yokoya, "Rationality of restricted re-quantization for efficient MPEG transcoding," *International Conference on Image Processing: Proceedings*, vol. 1, pp. 952–955, 2000.

[29] H. Sorial, W. E. Lynch, "Selective requantization for transcoding of MPEG compressed video," *IEEE International Conference on Multimedia and Expo, ICME 2000*, vol. 1, pp. 217–220, 2000.

[30] O. Werner, "Requantization for transcoding of MPEG-2 intraframes," *Image Processing, IEEE Transactions on*, vol. 8, p. 179191, Feb.1999.

[31] H. Sorial, W. E. Lynch, A. Vincent, "Estimating laplacian parameters of DCT coefficients for requantization in the transcoding of MPEG-2 video," *International Conference on Image Processing: Proceedings*, vol. 1, pp. 956–959, 2000.

[32] B. -U. Lee , "Modeling quantization error from quantised laplacian distribution," *El.Letters*, vol. 36, no. 15, pp. 1270–1271, July 2000.

[33] G. Zihua, O. C. Au, K. B. Letaief , "Parameter estimation for image/video transcoding," *ISCAS 2000 Geneva. The 2000 IEEE International Symposium on Circuits and Systems, 2000. Proceedings*, vol. 2, pp. 269 –272, 2000.

[34] "Information technology - generic coding of moving pictures and associated audio information - part 2: Video," *ISO/IEC 13818-2*, 1995.

[35] R. L. Lagendjik, E. D. Frimout, J. Biemond , "Low-complexity rate-distortion optimal transcoding of MPEG I-frames," *Signal Processing: Image Communication*, vol. 15, pp. 531–544, 2000.

[36] H.L. Wang, A.Divakaran, A. Vetro, S.F. Chang, H.F. Sun, "Survey of compressed-domain features used in audio-visual indexing and analysis," *Visual Communication and Image Representation*, vol. 14, pp. 150–183, Jun.2003.

[37] Y. Zhong, H.J. Zhang, A.K. Jain, "Automatic caption localization in compressed video," *Proc. IEEE Int. Conf. on Image Processing, Kobe, Japan*, Oct.1999.

[38] T.Y. Kim, J.H. Han, "Model-based discontinuity evaluation in the dct domain," *Signal Processing*, vol. 81, pp. 871–882, Apr.2001.

[39] C.S.Won, D.K. Park, I.Y. Na, S.-J.Yoo , "Efficient color feature extraction in compressed video," *Proc.SPIE Conf.on Storage and Retrieval for Image and Video Databases VII, San Jose, CA*, pp. 677 –686, 1999.

[40] K.T. Tan, M. Ghanbari, "Frequency domain measurement of blockiness in mpeg-2 coded video," *Proc. IEEE Int. Conf. on Image Processing*, vol. 3, pp. 977 –980, Sep 2000.

[41] R. W.-S. Chan, P. Goldsmith, "Modeling and validation of a psychovisually based image quality evaluator for dct-based compression," *Signal Processing-Image Communication*, vol. 17, pp. 485–49?, Jul.2002.

[42] Y.-H. Lee, J.-F. Yang, J.-F. Huang, "Perceptual activity measures computed from blocks in the transform domain," *Signal Processing*, vol. 82, pp. 693–707, Apr.2002.

[43] Y.Q, Liang, Y.-P. Tan, "A new content-based hybrid video transcoding method," *Proc. IEEE Int. Conf. on Image Processing*, vol. 1, pp. 429–432, Oct.2002.

[44] K.-L. Huang, Y.-S. Tung, J.-L. Wu, P.-K. Hsiao, H.-S. Chen, "A frame-based mpeg characteristics extraction tool and its application in video transcoding," *IEEE Trans. on Consumer Electronics*, vol. 48, pp. 522–532, Aug.2002.

[45] A. Puri, R. Aravind, "Motion-compensated video coding with adaptive perceptual quantization," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 1, no. 4, pp. 351–361, Dec.1991.

[46] R.V. Babu, K.R. Ramakrishnan, "Compressed domain motion segmentation for video object extraction," *ICASSP Proc.*, vol. 4, pp. IV–3788 –IV–3791, May 2002.

[47] S.Y. Kung, Y.-T. Lin, Y.-K. Chen, "Motion-based segmentation by principal singular vector (psv) clustering method," *ICASSP Proc.*, vol. 6, pp. 3410 –3413, May 1996.

[48] J.I. Khan, Z. Guo, W. Oh, "Motion based object tracking in mpeg-2 stream for perceptual region discriminating rate transcoding," *Proc. of the 9th ACM Int. Conf. on Multimedia*, pp. 572–576, Oct.2001.

[49] J. Wei, "Image segmentation using situational dct descriptors," *IEEE ISIP Proc.*, vol. 36, pp. 738–741, Oct.2001.

[50] S. Ji, H.W. Park, "Motion-based segmentation by principal singular vector (psv) clustering method," *Electronics Letters*, vol. 36, pp. 1769 –1770, Oct.2000.

[51] S. Ji, H.W. Park, "Region-based video segmentation using dct coefficients," *ICIP Proc.*, vol. 2, pp. 150–154, Oct.1999.

[52] S. Beucher and C. Lantuéjoul, "Use of watersheds in contour detection," *Workshop on Image Processing, Rennes, CCETT/IRISA, http://cmm.ensmp.fr/˜beucher/publi/watershed.pdf*, Sep 1979.

[53] M.L. Jamrozik, M.H. Hayes, "A compressed domain video object segmentation system," *ICIP Proc.*, vol. 1, pp. I–113 –I–116, Sep 2002.

[54] O. Sukmarg, K.R. Rao, "Fast object detection and segmentation in mpeg compressed domain," *TENCON 2000. Proc.*, vol. 3, pp. 364–368, Sep 2000.

[55] F. Moscheni, S. Bhattacharjee, M.Kunt, "Spatio-temporal segmentation based on region merging," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 897–915, Sep 1998.

[56] H.-L. Eng, K.-K. Ma, “Unsupervised image object segmentation over compressed domain,” *ICIP Proc.*, vol. 3, pp. 758–761, Oct 2000.

[57] H.-L. Eng, K.-K. Ma, “Spatiotemporal segmentation of moving video objects over mpeg compressed domain,” *IEEE ICME Proc.*, vol. 3, pp. 1531 –1534, Aug 2000.

[58] H.-L. Eng, K.-K. Ma, “Noise adaptive soft-switching median filter,” *IEEE Trans. on Image Processing*, vol. 10, pp. 242 –251, Feb 2001.

[59] H.-L. Eng, K.-K. Ma, “Bidirectional motion tracking for video indexing,” *IEEE Workshop on Multimedia Signal Processing*, pp. 153 –158, Sep.1999.

[60] N. Otsu, “A threshold selection method from grey-level histograms,” *IEEE Trans. Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 377–393, 1979.

בעבודה נעשות השוואות בין השיטות השונות, שהוזכרו לעיל, החל מפיענוח מלא של הסרט המקודד וקידודו מחדש בקצב הנמוך, וכלה בקידוד ישיר של סדרת המקור בקצב הנמוך בעזרת מקודד סטנדרטי TM5. הגישה הפשוטה, המבוססת על סיבוכיות המקרו-בלוקים, נתנה ביצועים טובים יותר מפיענוח מלא וקידוד מחדש כל עוד הורדת הקצב היתה לא יותר מאשר בפקטור של 2.5 (ביחס לקצב כניסה של 4Mbps בדוגמאות שנבחנו), אבל תמיד פחות טובים מאשר קידוד ישיר של סדרת המקור לאותו קצב. לעומת זאת, אופטימיזציה לגרנג׳יאנית לבחירת צעדי קוונוט אופטימליים כפי שהופיעה בסיפרות נותנת ברוב המקרים תוצאות טובות יותר מאשר קידוד של סדרת המקור לאותו קצב ע״י מקודד TM5. עם זאת, בחלק מהמקרים – בקצבים נמוכים, התוצאות כמעט זהות. ההרחבה המוצעת בעבודה זו נמצאה כנותנת את הביצועים הטובים ביותר מבין כל השיטות שנבחנו (מבחינת PSNR), במחיר עלייה בזמן ריצה של פי-3.3 ביחס לאופטימיזציה לגרנג׳יאנית רגילה.

הנושא האחרון שנבחן במחקר הנוכחי הינו שילוב של עקרונות ידועים של מערכת הראייה האנושית עם הגישות שפותחו בעבודה. למרות שבוצעו מחקרים רבים בתחום מערכת הראייה האנושית, לא קיים עדיין מודל מלא. יתרה מזאת, רוב הגישות שפותחו עד כה לא מתאימות למימוש במישור ה-DCT. הואיל והמחקר שלנו מתרכז בעבודה במישור ה-DCT, הוחלט לממש גישה פשוטה לצורך בניית מנגנון קידוד הנתמך ע״י מודל מערכת הראייה האנושית תוך שימוש בשיטות האופטימיזציה הלגרנג׳יאנית שהוצגו. בגישה המוצעת, התמונה הראשונה בקבוצה של תמונות (Group Of Pictures) מחולקת לשלושה סוגי איזורים: איזורים חלקים, גבולות ואיזורי מרקמים. החלוקה מתבצעת ברמה של בלוקים. מחשבים סכום הערכים המוחלטים של מקדמי AC ספיציפיים לקבלת מידע על רמת הפעילות של כל בלוק. אחר כך קובעים את הבלוקים בעלי אקטיביות גבוהה ע״י השוואה לסף אדפטיבי ובינריזציה, ומבדילים בין מרקמים לגבולות ע״י פעולות מורפולוגיות. בתמונות הבאות מבצעים עקיבה של החלוקה (סגמטציה) בהסתמך על וקטורי התנועה. בנוסף, מפעילים את תהליך החלוקה גם על תמונת ההפרשים ומוסיפים מידע לתמונת העקיבה במידת הצורך. בתהליך האופטימיזציה הלגרנג׳יאנית, השגיאה בכל בלוק משוקללת לפי סוג הבלוק וגודל תנועתו, הנתון ע״י וקטור התנועה. תוצאות של בדיקות לא פורמליות עם צופים אנושיים, ומדידות (באדיבות חברת אופטיבייס) בעזרת ציוד לבדיקת איכות וידאו – PQA200 של חברת Tektronix, מראות שבגישה זו משיגים אכן שיפור באיכות הסובייקטיבית של סרט הוידאו במוצא.

את המידע מכל המקרו-בלוקים שבתמונה, שיכולים להיות מקודדים עם צעדי קוונוט שונים, דבר שאינו אפשרי בגישות שערוך המבוססות על מודל לפלסיאני.

בנוסף, במהלך המחקר מומשה ונבחנה שיטה שהוצעה בספרות לבחירת צעדי הקוונוט האופטימליים ע״י אופטימיזציה לגרנג׳יאנית. בגישה זו הופכים את הבעיה של מינימיזצית השגיאה (עוות) עם אילוץ של הקצב הכולל לבעייה ללא אילוץ, ע״י שילוב של העוות והקצב בפונקציית מחיר אחת בעזרת פרמטר לגרנג׳יאני $\lambda\geq0$. במקרה זה ניתן לפתור בעיית מינימיזציה עבור כל מקרו-בלוק באופן בלתי תלוי ממקרו-בלוקים אחרים שבתמונה. אחרי ביצוע מינימיזציה עבור λ מסוים, בודקים את כמות הסיביות הכוללת הדרושה לקידוד התמונה. אם מספר הסיביות גדול מהרצוי, מגדילים את הפרמטר λ, אחרת מקטינים אותו. חוזרים על תהליך זה עד לקבלת כמות הסיביות הרצויה. כמו כן, נבחנה האפשרות של הפעלת גישות MSE ו-MAP יחד עם הגישה הלגרנג׳יאנית.

במהלך המחקר פותחה הרחבה מקורית לשיטה הלגרנג׳יאנית לעיל ע״י שינוי אופטימלי של האינדקסים של מקדמי ה-DCT המקוונטים. הגישה נותנת את הביצועים הטובים ביותר בין כל הגישות שהתפרסמו עד כה – במובן של PSNR. לצורך הורדת סיבוכיות הפתרון האופטימלי, פיתחנו מימוש יעיל של השיטה המבוסס על וריאנט של אלגורתם ויטרבי (Viterbi) לחיפוש המסלול האופטימלי בשבכה (Trellis). עפ״י גישה זו מציגים את המיקום של מקדם ספציפי בוקטור הסריקה (zig-zag scan vector) כדרגה (Stage) בשבכה, וכל ערך שנבחן באותו מקום מוגדר כמצב (State) באותה דרגה. על מנת לקבוע את המסלול האופטימלי למצב מסוים, בוחנים ב-trellis הרגיל את המסלולים מכל המצבים שבדרגה הקודמת. במקרה שלנו בוחנים מסלולים מכל הדרגות הקודמות, אך ממצב אחד ויחיד בכל דרגה. שוני זה נובע מהצורך לבחור באופן אופטימלי זוג run-level בקידוד ה-VLC. כמו כן, ניתן לשלב את גישת ה-MSE עם גישה זו. בעבודה הנוכחית הוצגו ונבחנו גם שינויים באלגוריתם המוצע להפחתת מספר החישובים: שינוי שאינו פוגע באופטימליות הפתרון, ושינויים שמביאים לפתרון תת-אופטימלי, אף עם הפחתה זעירה בלבד של הביצועים .

מנגנוני הורדת קצב וידאו מחולקות ברוב המקרים לשני חלקים: חלק האחראי על הקצאת סיביות לכל תמונה ברצף על סמך פרמטרים חיצוניים, ומנגנון להורדת הקצב של כל תמונה, שצריך לעמוד בדרישות של החלק הראשון. בעבודה אנו סוקרים מספר שיטות להקצאת סיביות לכל תמונה, אך לצורך מימוש בחרנו בגישה הפשוטה של הורדת כמות הסיביות לפי היחס של הקצבים הממוצעים בכניסה וביציאה של המערכת. גישה זאת אינה אופטימלית, אבל מונעת בעיות של גלישת חוצץ היציאה (Output Buffer).

כבסיס להשוואה מומש אלגוריתם פשוט ששייך למשפחה של אלגורתמים המבוססים על המושג "סיבוכיות (Complexity) של מקרו-בלוק", המוגדר כמכפלה של צעד הקוונוט של מקרו-בלוק בכמות הסיביות שנדרשו לקידודו. מושג הסיבוכיות בא להחליף את השונות (variance) של המקרו-בלוק המקובל במערכות קידוד וידאו, כמו למשל במקודד MPEG-2 הסטנדרטי TM5. ההנחה הבסיסית של השיטות המבוססות על סיבוכיות היא שכמות הסיביות הדרושה לקידוד של כל מקרו-בלוק, לאחר הפעלת אותו צעד קוונוט בכל המקרו-בלוקים שבתמונה, פרופורציונלית לסיבוכיות של כל מקרו-בלוק ומקרו-בלוק. בהגדרת הסיבוכיות משתמשים במידע שמפיקים מוידאו מקודד ללא פיענוח מלא, לעומת השונות, שמחשבים מהתמונות המקוריות – לפני קידודן.

במהלך המחקר מומשו שיטות להפחתת שגיאת הקוונוט מחדש (Requantization Error), הקרויות שיטת ה-MSE ושיטת ה-MAP. שיטת ה-MSE ממזערת את השגיאה הרבועית הממוצעת, בעוד ששיטת ה-MAP ממזערת את הסיכוי שהשגיאה הנ"ל תופיע. שתי הגישות מבוססות על ידיעה של פונקציות הפילוג של מקדמי ה-DCT או שערוכם מתוך המקדמים המקוונטים הידועים למערכת להורדת הקצב. בספרות דווחו התוצאות שניתנות להשגה בידיעת הפילוג של האות המקורי. כמו כן, פורסמו מספר מאמרים שהציגו שיטות שונות לשערוך פרמטר הפילוג מתוך הנתונים המקוונטים, בהנחה שהפילוג המקורי מתאים למודל לפלסיאני. לאחר מימוש ובדיקת הביצועים התגלה ששימוש באינטרפולציה לינארית לצורך שערוך פונקצית הפילוג של מקדמי ה-DCT נותן תוצאות טובות יותר מאשר השימוש במודל לפלסיאני עבור השימוש הנדון. כמו כן, אינטרפולציה לינארית מאפשרת לנצל

תקציר

בשירותי מולטימדיה משתמשים כיום בכמות גדולה של חומר וידאו דחוס בפורמטים שונים לצרכי איכסון ושידור. נכון להיום, תקן MPEG-2 הינו התקן העיקרי לקידוד וידאו. הגמישות של התקן מאפשר שימוש במגוון רב של שימושים, כולל וידאו ספרתי, למידה מרחוק, וידאו on demand, וכו׳. בקידוד וידאו, לצורך שידורו ברשת, נתוני ערוץ השידור חייבים להיות מוגדרים מראש. מצד שני, רשתות מחשב מורכבות מאוסף רב של רשתות שונות המחוברות יחדיו. לכן פרמטרי שידור, כמו פורמטי וידאו נתמכים ורוחב סרט זמין, עלולים להשתנות בצורה משמעותית כתלות בזמן ובמשתמש. הבעיה מתעוררת במיוחד כאשר יש צורך להעביר אותו תוכן למספר יעדים. הפתרון הנפוץ ביותר כיום הוא להתחיל מסרט וידאו באיכות גבוהה, שניתן לקבלו בעזרת קידוד בזמן אמת או אחזור מאכסון, ולהתאימו לדרישות רוחב סרט של יעד ספציפי ע״י הורדת הקצב (Transrating).

ישנן כיום שיטות רבות להורדת הקצב של וידאו מקודד. המטרה של הורדת הקצב הינה להשיג קצב רצוי תוך שמירה על האיכות המקסימלית האפשרית של סרט הוידאו במוצא. הפתרון הנאיבי של פיענוח מלא וקידוד מחדש של רצף התמונות המפוענח אינו מוצלח בגלל הסיבוכיות הגבוהה שלו, הנובעת בעיקר משערוך מחודש של וקטורי התנועה. כמו כן, בפתרון זה יש ירידה באיכות המתקבלת של הוידאו עקב אי-דיוקים בביצוע התמרות DCT/IDCT, שגיאות של קוונוט מחדש (Requantization) ואי-אופטימליות של שערוך וקטורי התנועה הנובעת משימוש בסדרת הוידאו המשוחזרת כסדרת מקור. כיום, רוב השיטות להורדת הקצב פועלות במישור ה-DCT ומנצלות את ידיעת ההחלטות של המקודד הראשוני לצורך שיפור ביצועיהן. השיטה הנותנת את האיכות הטובה ביותר, במובן PSNR, מבין שאר השיטות שפותחו קודם, משתמשת באופטימיזציה לגרנג׳יאנית למציאת צעדי קוונוט חדשים אופטימליים.

במחקר הנוכחי מומשו ונבחנו גישות שונות להורדת הקצב של וידאו מקודד ע״י קוונוט מחדש ומוצעת שיטה חדשה שמביאה לביצועים טובים יותר (מבחינת ה-PSNR). זאת בנוסף להתייחסות לתכונות מערכת הראייה האנושית כדי לשפר את האיכות הויזואלית של מערכות להורדת קצב.

רשימת ציורים

רשימת טבלאות

תוכן עניינים

המחקר נעשה בהנחיית פרופ' דוד מלאך בפקולטה להנדסת חשמל.

המחקר מומן חלקית ע"י קונסורציום STRIMM
במסגרת תוכנית מגנ"ט של לשכת המדען הראשי בתמ"ת, באמצעות
מוסד שמואל נאמן למחקר מתקדם במדע וטכנולוגיה.

אני מודה לטכניון על התמיכה הכספית הנדיבה בהשתלמותי.

הנני אסיר תודה למנחה שלי, פרופסור דוד מלאך , על הנחייתו
המסורה ותמיכתו הבלתי מסויגת לאורך כל תקופת המחקר.

אני מודה למעבדה לעיבוד אותות, ובפרט לנימרוד פלג, איתי סגל,
זיו רביזנוביץ ואריק יאבילביץ, על עזרתם בפיתוח מערכת לניתוח
וידאו מקודד . הנני מודה גם לחברת אופטיבייס, ובפרט לאילן
דניאל, על העמדת הציוד לבדיקת איכות וידאו סובייקטיבית לרשותי
והעזרה בביצוע הבדיקות.

הורדת קצב של וידאו מקודד ע"י קוונוט מחדש אופטימלי

חיבור על מחקר

לשם מילוי חלקי של הדרישות
לקבלת התואר מגיסטר למדעים
בהנדסת חשמל

מיכאל לברנטייב

הוגש לסנט הטכניון - מכון טכנולוגי לישראל

ניסן, תשס"ד   חיפה   אפריל 2004

הורדת קצב של וידאו מקודד ע"י קוונוט מחדש אופטימלי

מיכאל לברנטייב