UNIVERSIDADE DA BEIRA INTERIOR
Covilhã | Portugal

Robust video coder solution for wireless streaming: Applications in Gaussian channels.

Ângelo Miguel Arrifano

Submitted to University of Beira Interior in candidature for the degree of
Master of Computing and Intelligent Systems

Supervised by Maria Manuela Pereira

Departamento de Informática
University of Beira Interior
Covilhã, Portugal
October 2010

Acknowledgements

I want to thank my family and friends, in special to my father Ângelo Arrifano and my mother Georgina Arrifano for their unconditional support. A big hug to my uncle Pedro Pires that offered me a MSX Basicat my early 12 years old, where I started by myself learning computer programming. A tiny hug for his son David, the one I have the pleasure to be the godfather.

My greetings to the holography group of the secondary school "Escola Secundária Campos Melo", and its leaders Prof. Pedro Pombo and Prof. Rosa Simões, as well the other persons involved, for awakening my interest in Science.

I would also like to give a big hug to my friends at the I3S laboratory, in France.

At last, but not least, I want to express my gratitude to the teachers Marc Antonini e Manuela Pereira for their support. My acknowledgments to the teacher Mário Freire for always making himself available to help.

Abstract

With the technological progress in wireless communications seen in the past decade, the miniaturization of personal computers was imminent. Due to the limited availability of resources in these small devices, it has been preferable to stream the media over widely deployed networks like the Internet. However, the conventional protocols used in physical and data-link layers are not adequate for reliable video streaming over noisy wireless channels. There are several popular and well-studied mechanisms for addressing this problem, one of them being Multiple-Description-Coding. However, proposed solutions are too specialized, focusing the coding of either motion or spatial information; thus failing to address the whole problem, that is - the robust video coding.

In this thesis a novel MDC video coder is presented, which was developed during an internship at the I3S laboratory - France. The full coding scheme is capable of robust transmission of Motion-Vectors and wavelet-subband information over noisy wireless channels. The former is accomplished by using a MAP-based MD-decoding algorithm available in literature, while the robust transmission of wavelet-subbands is achieved using a state-of-the-art registry-based JPEG-2000 MDC. In order to efficiently balance MV information between multiple descriptions, a novel R/D-optimizing MD bit-allocation scheme is presented. As it is also important to efficiently distribute bits between subband and motion information, a global subband/motion-vector bit-allocation technique found in literature was adopted and improved. Indeed, this thesis would not be complete without the presentation of produced streams as well as of a set of backing scientific results.

keywords: multiple description coding, video coding with temporal prediction, motion compensated video coding, wireless channel, noisy channel, gaussian channels, 3g, unreliable channels, robust video coding, temporal wavelet transform, motion compensation, subband coding, motion vector coding, maximum a posteriori estimation, jpeg 2000, registry based jpeg 2000, bit allocation, lagrangian allocation.

Contents

List of Figures

Acronyms

- ARQ - Automatic Repeat-reQuest.
- AWGN - Additive White Gaussian Noise.
- BPSK - Binary Phase-Shift Keying.
- BSC - Binary-Symmetric-Channel.
- BsC - Bachelor of Science.
- FEC - Forward Error Correction.
- GOP - Group of Pictures (context of temporal coding).
- IP - Internet Protocol.
- ISO - International Organization for Standardization.
- JPEG - Joint Photographic Experts Group.
- JPEG-2000 -The image compression standard created in 2000.
- MAP - Maximum-a-Posteriori probability estimation.
- MCS - Most Common Sign.
- MDSQ - Multiple-Description-Scalar-Quantization.
- MDTC - Multiple-Description-Transform-Coding.
- MCWT - Motion-Compensated WT.
- MDC - Multiple Description Coding.
- MD - Multiple-Description.
- WT - Wavelet-Transform.
- MsC - Master of Science.
- MV - Motion-Vector.
- OSI - The Open Systems Interconnection model from ISO.

- PSNR - Peak SNR.
- QoS - Quality of Service.
- SB - Sub-Band (wavelet).
- SNR - Signal-to-Noise Ratio.
- SDC - Single-Description-Coding.
- UMPC - Ultra-Mobile Personal Computer.

Chapter 1

Introduction

With the technological progress in wireless communications seen in the past decade, the miniaturization of personal computers was imminent. From early handhelds to current bleeding-edge Ultra-Mobile-Personal-Computers (UMPC), all shared the need to playback media content on-demand. Moreover, due to the limited availability of resources on such devices, it has not been convenient to store media on them. A preferable way is then media streaming, which is already widely deployed in the major IP network - the Internet. However, wirelessly streaming media over IP is prone to problems common to all wireless channels, such as round-trip and retransmission latencies.

The Multiple-Description-Coding (MDC) problem, as first put by Gersho, Witsenhausen, Wolf, Wyner, Ziv and Ozarow at the 1979 Shannon Theory Workshop [4], consists in describing an information source by two separate descriptions in such manner that:

- when one of the descriptions is decoded separately, it allows the source reconstruction with acceptable quality;
- when both descriptions are decoded jointly, they provide a high-quality reconstruction of the source.

The aforementioned coding scheme, depicted in figure 1.1, has seen a rise in interest by the source-coding community due to its advantages for video coding and transmission over noisy channels, compared to packet retransmission and Forward-Correcting-Codes (FECs). Indeed, in the context of video streaming it is unacceptable to rely on packet retransmission in the case of a transmission error; during playback it introduces undesirable latency between video frames that ruins the perception of motion. There were several attempts to reduce packet retransmission latency by means of Quality-of-Service (QoS) and traffic shaping, but these only work for small amounts of channel errors. On the other hand, FECs are a useful mechanism for keeping data integrity along transmission; it works by sending enough redundant bits merged with the actual data to the decoder, providing the ability to correct data on-the-fly avoiding packet retransmission [5]. However, FEC implementations are mostly useful to protect small payloads and must be parameterized with care [5].

Starting from a few years ago and in the sequence of the Bachelor (BsC) and during the Master (MsC) programs at the University of Beira Interior (UBI) - Portugal, the author was first introduced to and then further explored MDC as a potential solution for robust image and video transmission

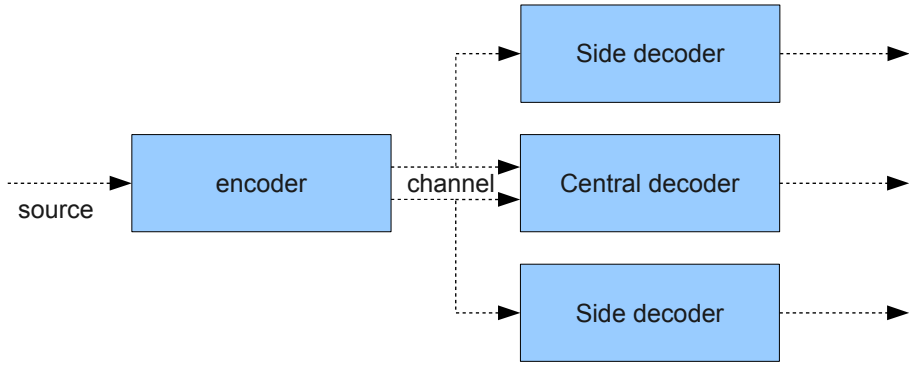


Figure 1.1: General multi-description-coding scheme.

over wireless channels. Based on a Motion-Compensated-Wavelet-Transform (MCWT) video coder, the objective was to exploit the then state-of-the-art JPEG-2000 for spatial MDC. This work was made under supervision of Manuela Pereira from the same University and in cooperation with the I3S laboratory through Marc Antonini, which led to the publishing of two scientific papers[6][1]. However, the developed video coder[1] unreasonably required that Motion-Vectors (MVs) were sent over a dedicated noiseless channel[1], which is not suitable for most wireless communication systems.

Then, in the context of a internship at I3S laboratory in France, the registry-based JPEG-2000 MCWT MD video coder [1] was complemented with a state-of-the-art MDC solution to address the required robust transmission of Motion-Vectors over wireless links. Since balancing the MV bits between descriptions is indeed a critical task, a novel R/D-optimizing MD bit-allocation scheme for motion vectors is also presented. These contributions are indeed fundamental to successfully achieve the goal of robust video coding. As it is also important to efficiently distribute bits between subband and motion information, the global subband/motion-vector bit-allocation technique found in [7] was adopted and improved.

In Chapter 2 a valuable background on past and current robust video coding solutions is provided. The MD robust video coder is introduced in Chapter 3 with a explanation of its overall functionality, including the improvements in the global bit-allocation scheme[7]. Chapter 4 presents the MCWT Subband MDC scheme based on JPEG-2000 [1], along some related results. Chapter 5 then focuses two more improvements achieved in the robust video coder: the MV MD coder and the MV MD bit-allocation schemes. Chapter 6 highlights the overall video coder results and finally, in Chapter 7 the thesis is concluded.

Chapter 2

State of the Art

2.1 The Robust Video Coding

The research in source coding techniques led to the development of important video coding standards during the past twenty years - the MPEG-1, MPEG-2 and latest MPEG-4 from ISO; the H.263 and latest H.264 from ITU. However, error resilience was not the main objective of these coding standards, with the solutions to that problem being often addressed in optional annexes (as in H.263). The coding standard H.264/MPEG-4 AVC goes beyond that and specifies some error-resilience features for error isolation, data recovery and error concealment [8]. Still, the error-resilience features available in coding standards are not enough to provide a minimum acceptable video quality in presence of wireless channel errors [9] [10].

There is, however, an extensive set of publications that try to solve this problem by using several different techniques; some of them care to keep compatibility with current standards, while others propose group-up designs to achieve better error-resilience. Compatibility with the coding standards is an advantage because it eases deployment of the solutions. Error-resilience techniques can be split into Single-Description-Coding (SDC) and Multiple-Description-Coding (MDC). SDC is by far the most common way of source coding, being used in all of the coding standards. However, because good error-resilience is difficult and complex to achieve in SDC, the MDC has seen a rise of interest by the source and channel coding communities.

2.2 The Multiple-Description-Coding

The Multiple-Description-Coding (MDC) problem, as first put by Gersho, Witsenhausen, Wolf, Wyner, Ziv and Ozarow at the 1979 Shannon Theory Workshop [4], consists in describing an information source by two separate descriptions in such manner that:

- when one of the descriptions is decoded separately, it allows the source reconstruction with an minimum acceptable quality;
- when both descriptions are decoded jointly, they provide a high-quality reconstruction of the source.

In 1980 the first theoretical results of the achievable rates of MDC in a Binary-Symmetric-Channel (BSC) were published in [11][12][13][14] by Witsenhausen, Ozarow, Wolf, Wyner, Ziv, El Gamal and Cover; and later in 1981, one of the first practical MDC applications is presented in [15] where ADPCM coded speech samples are split even/odd.

After almost 30 years, the MDC is still under wide research with most of its applications in robust image and video coding. With the increasing interest in Peer-to-Peer (P2P) multimedia streaming and with the fundamental architecture of P2P using Multiple-Descriptions, MDC is also a promising solution for robust P2P video streaming [16] [17]. Among most popular MDC schemes is the MD-Scalar-Quantization (MDSQ) pioneered by Vaishampayan in [18], the MD-Transform-Coding (MDTC) pioneered by Wang and Orchard in [19] and MDC based in FEC pioneered by Puri *et all.* in [20].

Chapter 3

The robust video coder

3.1 Overview

In order to increase the robustness of the JPEG-2000 based MC-WT MD video coder[1] a scheme visible in figures 3.1 and 3.2 is proposed. This updated framework includes a slightly modified version of the model-based wavelet-subband/motion-vector bit-allocation algorithm proposed by Marie André *et all.* in [7]. The algorithm, detailed in section 3.2, allows the rate-control between motion-vectors R_v and wavelet subbands R_c in such a way that overall distortion is minimized. Optimal rates for subbands and MVs encoding (R_c and R_v) are then provided to their respective Multiple-Description (MD) bit-allocation algorithms (fig. 3.1) to efficiently dispatch bits between available descriptions, according to channel characteristics such as Bit-Error-Rate (BER).

The MD wavelet subband encoding is accomplished by first performing a MD bit-allocation algorithm that is the implementation of the proposal presented in [21] by Manuela *et all.* This algorithm provides a set of rates $R_{i,j}$, where $i = 1..\#SBs$ and $j = \{1, 2\}$, that minimizes the central subband reconstruction. Afterwards, each wavelet subband description is encoded with the registry-based MD JPEG-2000 [1] using the optimally allocated rates $R_{i,j}$ (fig. 3.1).

The MD motion-vector encoding is done by first performing an optimizing MD bit-allocation algorithm that distributes bits between MV descriptions in a way that minimizes the central MV reconstruction. For this MV MD bit-allocation, the author studied and developed a new scheme which is fully explained in Chapter 5. First, two descriptions are created where each contains a copy of all the motion vectors. Then, each description is split into two disjoint groups which are quantized using rates Rv_{gj} provided by the bit-allocation, being $j \in \{1, 2\}$ the description and $g \in \{1, 2\}$ the vector group. Therefore each description is quantized using an independent rate $Rv_j = Rv_{1j} + Rv_{2j}$.

A carrier signal with Binary Phase-Shift Keying (BPSK) modulation subject to Additive White Gaussian Noise (AWGN) is considered, since it provides a simple but yet good model of a generic wireless channel. To overcome packet retransmission the video coder operates directly on physical and data-link layers of OSI model, that is, it is responsible for modulation and error control. In this way, wavelet-subbands and motion-vectors descriptions are modulated and transmitted to the decoder through disjoint channel paths.

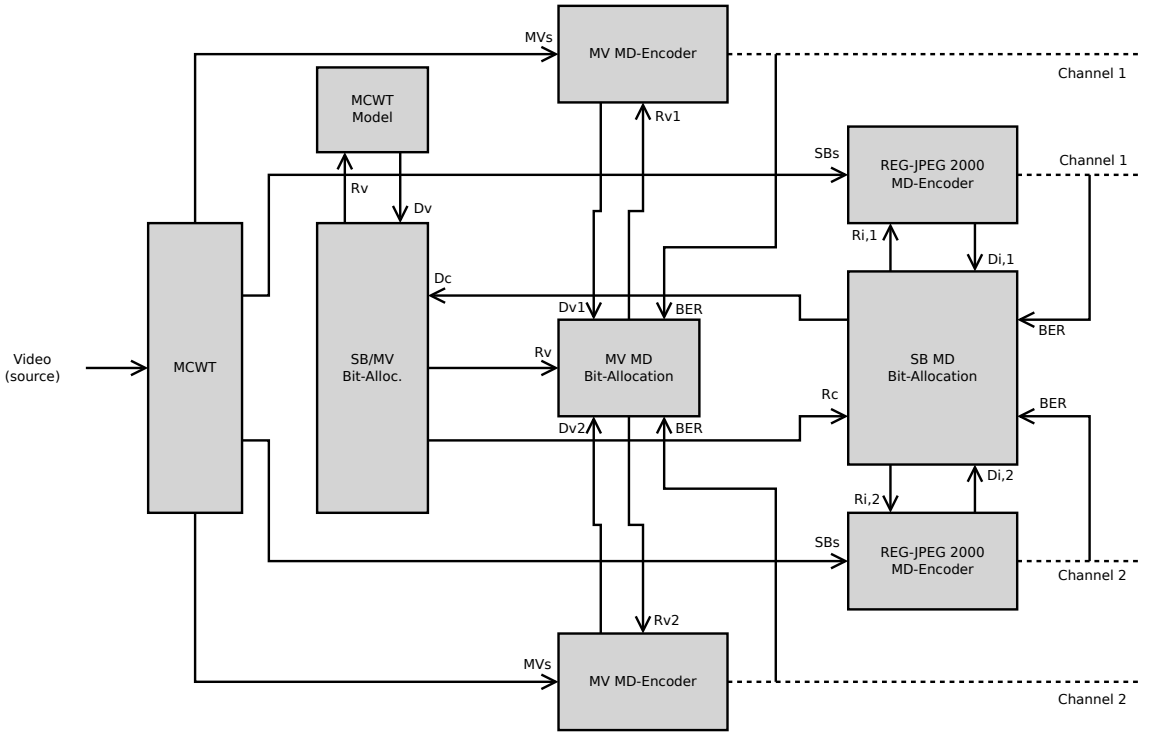


Figure 3.1: The proposed framework for robust video encoding, featuring the MV MDC and MV MD bit-allocation schemes.

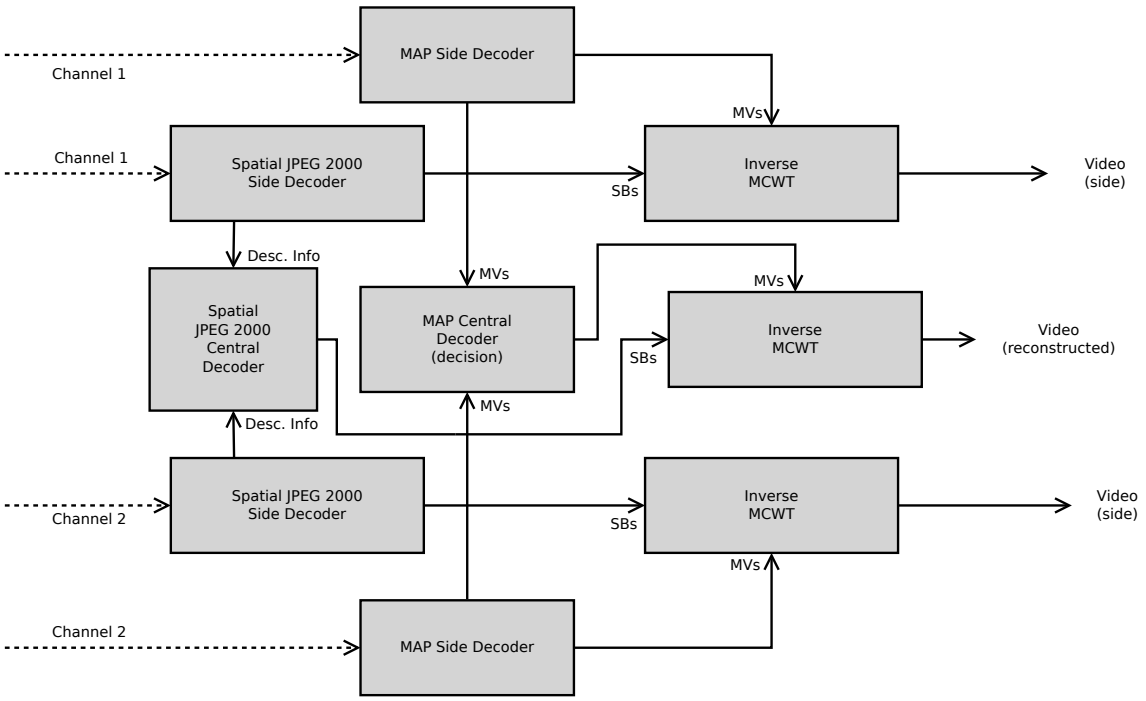


Figure 3.2: The proposed framework for robust video decoding.

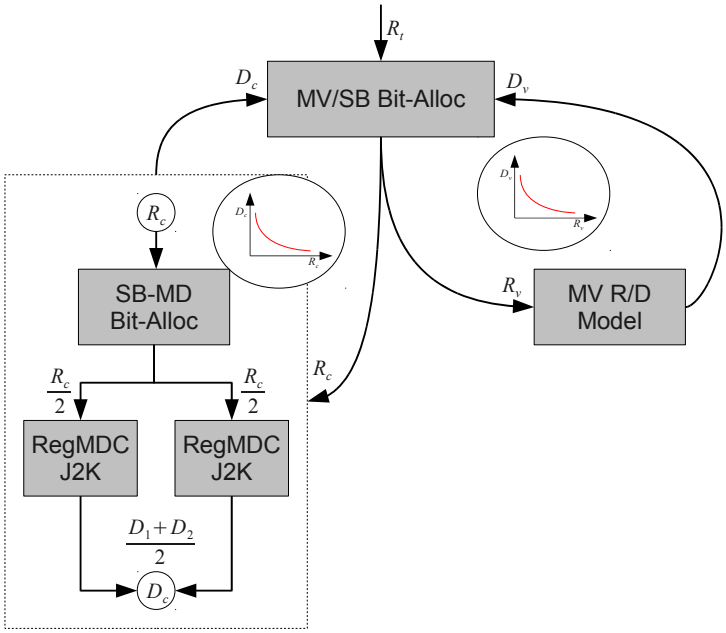


Figure 3.3: The proposed rate/distortion scheme for bit-allocation between motion information and SBs.

In the decoder, the main improvement lies in the MD-decoding of MVs as evidenced through the comparison of figures 4.1 and 3.2. The MVs that were quantized, before transmission, with different quantization steps are optimally demodulated and dequantized using a MAP estimator [2]. A study of this MAP estimator is included in Chapter 5. The JPEG-2000 based MD decoder does not take any advantage in dealing with modulated data, therefore wavelet-subbands are hard-demodulated before decoding.

3.2 MV/SB Bit-Allocation

The model-based bit-allocation between Subbands (SBs) and motion information presented in [7] was integrated in the context of current framework (3.1). Given a total rate R_t, the objective is to allocate the rate for motion-vectors R_v and wavelet-subbands R_c, so that $R_t = R_c + R_v$ and the overall distortion $D(R_v, R_c)$ is minimized.

The proposed global bit-allocation is an optimized implementation of the algorithm published in [7]. First, a $D(R_v)$ curve 3.3 is computed by spline-interpolation using M reference points; this gives the MCWT output distortion D_v caused by MV quantization that is controlled by rate R_v (fig. 3.4). Each reference point is calculated using model [7] instead of actually performing the MCWT, which

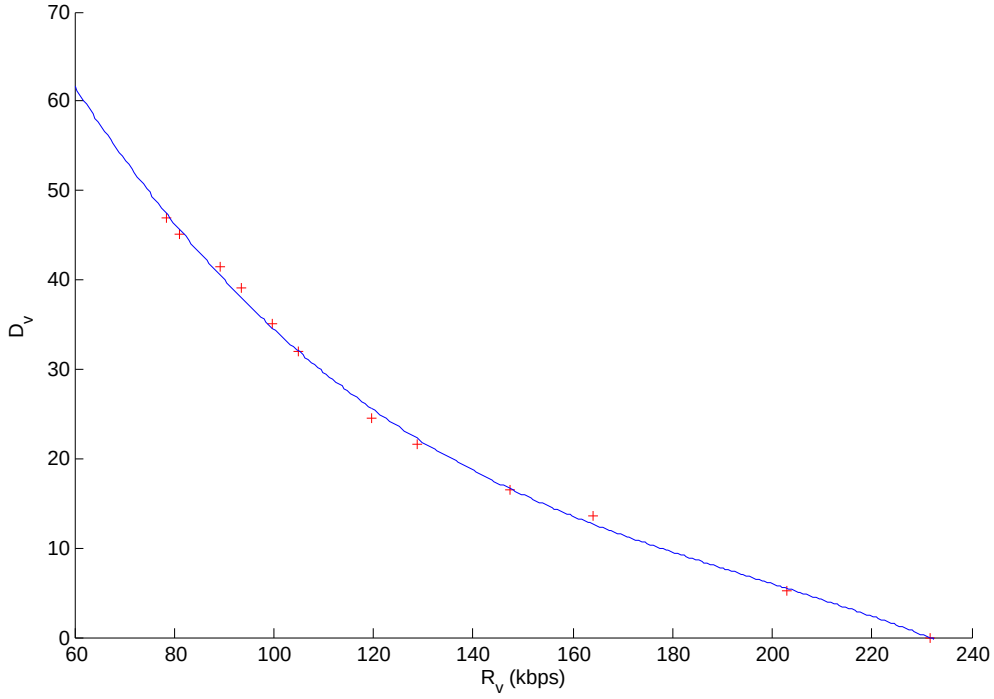


Figure 3.4: MCWT output distortion D_v due to MV quantization controlled by rate R_v. "Foreman" sequence, two decomposition-levels.

is of faster execution:

$$D_v = \frac{1}{2K} \sum_{n=0}^{N-1} \sum_{k=0}^{\frac{K}{2^{N-n}}-1} \Big[Pn(x_{2^{N-n}k}) - \Gamma_{x_{2^{N-n}k}}(\eta_{B_{2^{N-n}k+2^{N-n-1}}}) + Pn(x_{2^{N-n}k+2^{N-n}}) - \Gamma_{x_{2^{N-n}k+2^{N-n}}}(\eta_{F_{2^{N-n}k+2^{N-n-1}}}) \Big] \tag{3.1}$$

K is the size of the sequence, N the number of decomposition levels, $Pn(x)$ is the power of image x and $\Gamma_x(y)$ is the autocorrelation function of image x at lag y. η_{B_x} and η_{F_x} are the quantization errors in the backward and forward motion vectors.

Given a value for lambda, for each of the M test R_c points an associated $R_v^*(\lambda, R_c)$ [7] is calculated by minimizing the criterion $J_\lambda(R_v, R_c)$ [7] in function of R_v:

$$J_\lambda(R_v, R_c) = D(R_v, R_c) + \lambda(R_v + R_c + R_t) \tag{3.2}$$

The total distortion $D(R_v, R_c)$ is the distortion produced by motion compensation $D(R_v)$ plus the distortion of subband quantization $D_c(R_c)$. Since the video framework performs more than a simple quantization in the wavelet subbands, the rate/distortion $D_c(R_c)$ is actually measured through an experimental coding phase. This experimental coding phase is represented as the dashed box in figure 3.3. Inside this "black-box", the total subband rate R_c is first allocated to each description by means of the MD Bit-Allocation algorithm published in [21]. Each temporal subband is then encoded using the Registry-based MD-JPEG-2000 [1] using allocated rates $R_{i,j}, i = 1..\#SBs, j = \{1, 2\}$; and the

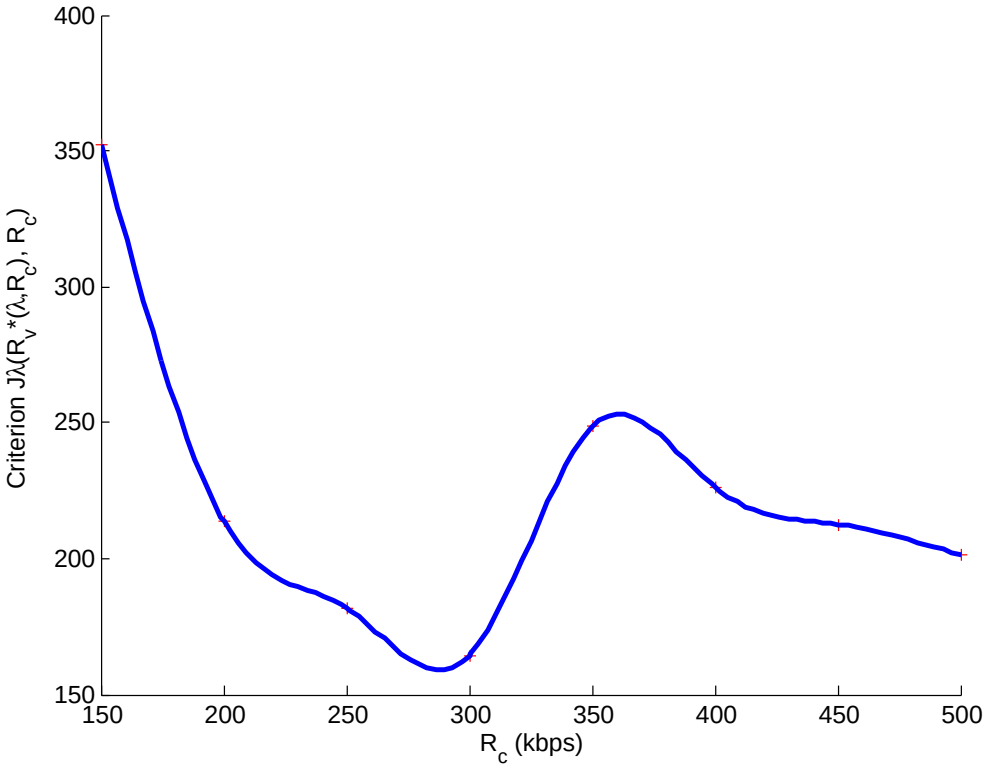


Figure 3.5: Criterion $J_\lambda(R_v^*(\lambda, R_c), R_c)$. "Foreman" sequence, $R_t = 500$ kbps, two decomposition-levels, $\lambda \approx 0.015774$. Minimum with $R_c \approx 288.110680$

output distortions

$$\begin{aligned} D_j &= \frac{1}{\#SBs} \sum_{i=1}^{\#SBs} D_{i,j}, j = 1, 2 \\ D_c &= \frac{D_1 + D_2}{2} \end{aligned} \tag{3.3}$$

measured. Proceeding in this way, by calculating D_c as the average distortion between descriptions, it allows easily minimization of the central distortion D_0 in the MV/SB Bit-Allocation procedure.

The M $R_v^*(\lambda, R_c)$ points are then spline-interpolated (fig 3.5) to find the R_c^* that minimizes the criterion 3.2. Having found the optimal R_c^* and R_v^*, the Bit-Allocation algorithm must ensure that

$$R_c^* + R_v^* \approx R_t \tag{3.4}$$

To meet this constraint, in [7] it is proposed that λ is tuned by dichotomy. However, in this implementation, the minimization of criterion 3.2 is performed at 5 linearly-spaced λ and then the results are interpolated (fig. 3.6). Therefore, it is possible to quickly estimate the value of λ that meets the constraint 3.4.

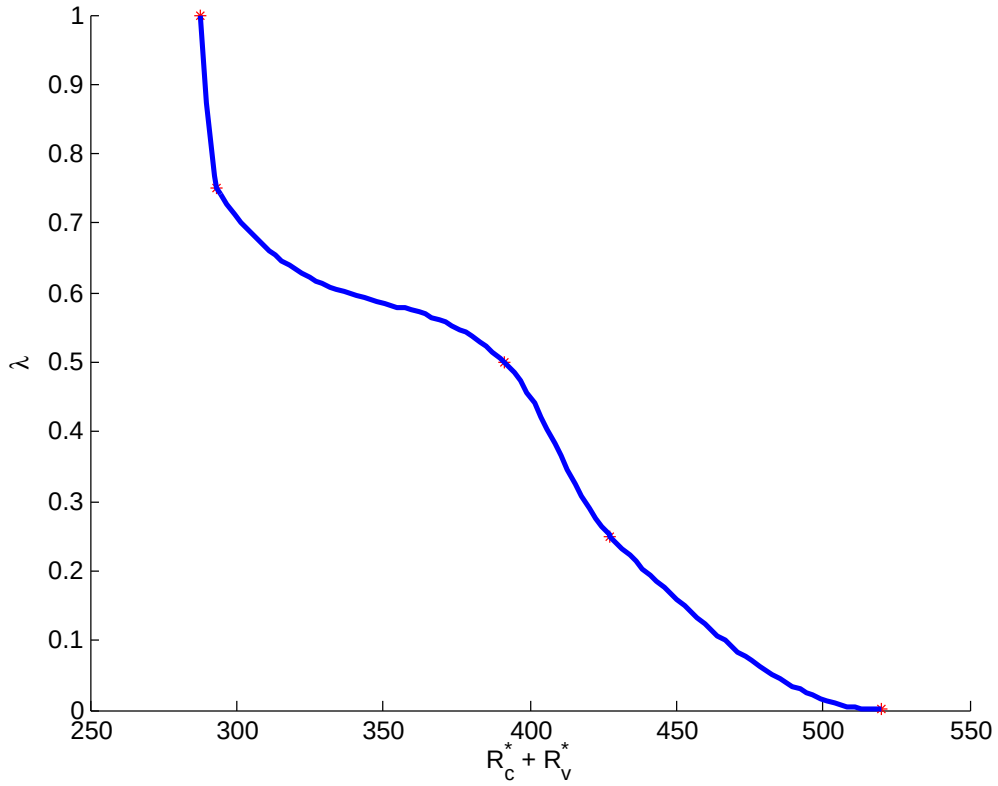


Figure 3.6: The minimization result $R_c^* + R_v^*$ of criterion 3.2 for different λ. For $R_t = 500$ the estimated $\lambda \approx 0.015774$, where $R_c^* = 288.110680$ and $R_v^* = 231.599954$.

Chapter 4

Multiple-Description-Coding of Wavelet-Subbands

Due to the success of JPEG-2000, there are several proposals of MDC using JPEG-2000 in literature. In [22] the authors present a rate-distortion-based MDC compatible with JPEG-2000 that is enhanced in [23] with a MD quantization step. In [24], a prediction-compensated MDC using filter banks is presented, which is compatible with JPEG-2000. However, they do not provide a scheme for MD video coding and that is the main reason why the use of MDC makes sense. They also share a common particularity - the aim to optimize redundancy-allocation between descriptions. The proposed MDC method, also published in [1], goes beyond that by relying on error-detection capabilities to provide an optimal central reconstruction in the absence of a full description or just a few bits.

In section 4.1, it is presented a method for MD image coding using JPEG-2000. It allows the encoding of descriptions provided with key-registers for error-detection, which are still JPEG-2000 compatible. The multiple-description JPEG-2000 decoder is then capable to precisely detect transmission errors and efficiently choose between available description-information using those registers, achieved by a clever exploitation of the Embedded Block Coding with Optimized Truncation (EBCOT) system. In section 4.2, the modified JPEG-2000 MD-coder is integrated as a spatial MD-coder (for encoding motion-compensated temporal wavelet subbands) in the video framework[25]. The presented video coder scheme (fig. 4.1) is inherently[21] capable of bit-allocation of JPEG-2000 subband data between multiple descriptions, being then able to adjust description redundancy according to channel characteristics. Some experimental results are presented in the last section 4.3 within this chapter.

4.1 MQ-Coder Registry-Based MDC

4.1.1 The MDC Scheme

The proposed MD image coding scheme is represented in fig. 4.2. A source signal (image) is encoded twice with the JPEG-2000 encoder presented in [6], to produce two codestreams (descriptions). The

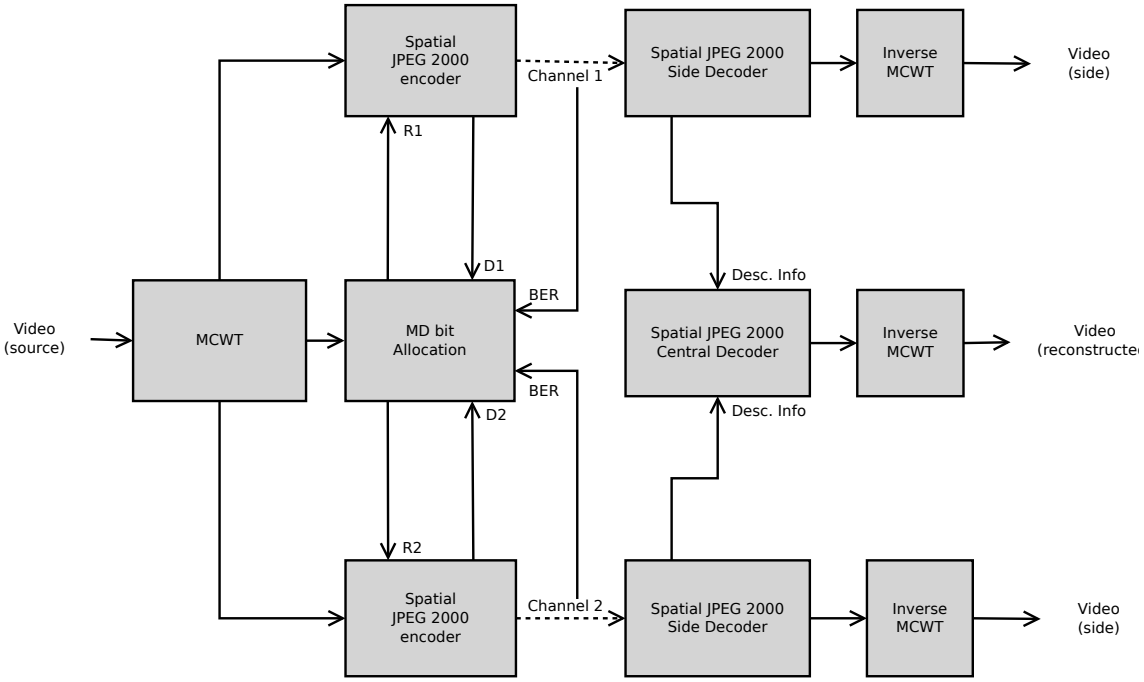


Figure 4.1: The JPEG-2000 based MCWT MD video coder [1].

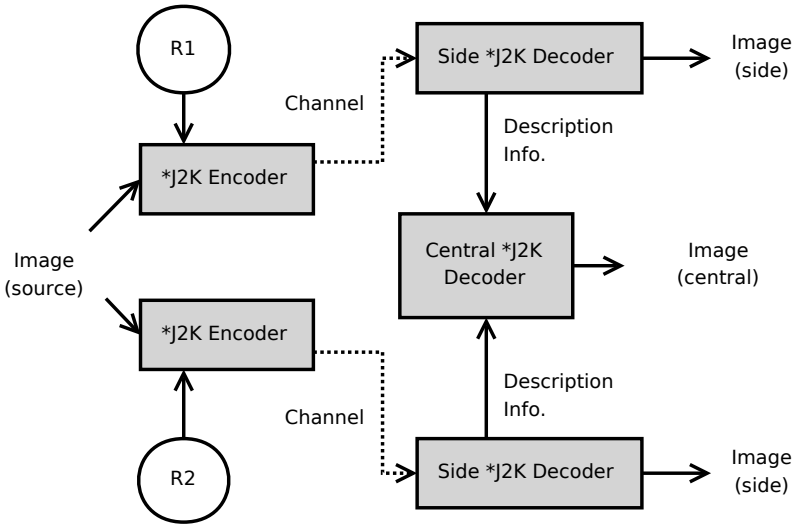


Figure 4.2: Side/Central JPEG-2000 MDC scheme for two transmission channels.

rate of each description can be adjusted using the standard JPEG-2000 rate (R_1 and R_2 on fig. 4.2) control algorithm. In the case of a video coder, the input rate can be acquired from the MD bit-allocation algorithm presented in [25].

Each description is then sent to the side decoder over the corresponding transmission channel. The specialized JPEG-2000 side decoder, carefully explained in section 4.1.2, receives each description and proceeds with full decoding, producing two distinct outputs: the decoded image and the description-information. The decoded image is simply the output of the standard JPEG 2000 decoder with the description received as is. The description-information is a bitstream containing key-information about the description decoding process and includes the following information:

- MQ-Encoder registers saved during the encoding process.
- MQ-Decoder registers gathered from a side decoding process.
- Complete J2K codestream (or pointers to its location in the description).

Description-information is then read by our specialized JPEG-2000 central decoder (fig. 4.2), allowing it to make key decisions about the best information to use to optimally reconstruct the original image. The above explanation is based on two-channel MDC only for the sake of simplicity, the proposed method is scalable to any given set of transmission channels.

4.1.2 The Specialized JPEG 2000 Side/Central Decoders

Given that JPEG-2000 encoder (fig. 4.2) produces codestreams provided with some special registers as described in [6], the side decoder can successfully detect which JPEG-2000 segments are corrupted by comparing the encoder-saved special registry with the corresponding decoder registry, very much like explained in [6].

The MQ-encoder is ran with the *RESTART* [26] mode switch enabled, which forces the restart of the MQ-coder at the beginning of each coding pass. Besides the obvious error-resilience provided by this encoding mode, it is specifically needed to flush the arithmetic coder at the end of every coding pass and thus making them valid truncation points. The MQ-coder A [27] register is then saved at the end of each coding pass and later inserted into the J2K codestream. During the decoding process, the MQ-encoder-saved registers are compared with the MQ-decoder A registers at every segment. In table 4.3 we can see an example of registry checks for both side decoders using different rates R_1 and R_2. Registers are presented in the form of `[xxxx/yyyy]`, where `xxxx` is the decoder registry and `yyyy` is the encoder-saved registry. If they do not match, then it is highly probable that the JPEG-2000 segment is corrupted [6].

When the central decoder runs, each available description-information (fig. 4.2) contains all the information needed to decide which description segment to use so that the reconstructed image SNR is maximized - for example, the segment status (corrupt/non-corrupt) and segment data-length.

4.1.3 JPEG 2000 Codestream Compatibility

Since the produced codestreams are still JPEG 2000 compliant [6], standard JPEG 2000 decoders can be used to decode a single-description, which provides similar results to the single-description coding

Seg.	Desc_1	Desc_2	Central decision
00	[a834/a834]	[a834/a834]	Choose any
01	[9807/9807]	[9807/9807]	Choose any
02	[ac02/ac02]	[ac02/ac02]	Choose any
03	[a807/a807]	[a807/a807]	Choose any
04	[bffb/ae93]	[ae93/ae93]	Choose d_2
05	[d006/d006]	[d006/d006]	Choose any
06	[f806/83fe]	[83fe/83fe]	Choose d_2
07	[a202/d804]	[d804/d804]	Choose d_2
08	[9002/fc04]	[fc04/fc04]	Choose d_2
09	[a202/8e04]	[8e04/8e04]	Choose d_2
10	[a802/c805]	[c805/c805]	Choose d_2
11	[e008/a404]	[a404/a404]	Choose d_2
12	[cffe/c3ff]	[c3ff/c3ff]	Choose d_2
13	[f002/f002]	[f002/f002]	Choose any
14	[ec02/c804]	[9c07/c804]	Discard
15	[8000/c002]	[c002/c002]	Choose d_2
16	[900b/b802]	[b802/b802]	Choose d_2
17	[f008/ae02]	[d80b/ae02]	Discard
18	[8000/8000]	[8000/8000]	Choose any
19	[9004/c004]		Discard
20	[8c03/ac02]		Discard
21	[8000/8000]		Choose d_1

Figure 4.3: MQ-Coder registry-based J2K segment check: segments are chosen by the central decoder according to their [current/saved] register values. The descriptions were encoded at different rates.

(SDC) mode. However, the subsequent use of a side and then of a central decoder to decode a single-description still benefits from a minimum of error-detection capabilities, since corrupted segments are always discarded.

4.1.4 MQ-coder Register Overhead

The register overhead is 16bit [6] per segment. The total number of segments in a codestream directly influences the error-resilience level and overhead. Higher error-resilience requires a higher number of smaller segments, so a higher number of registers is used. The best error-resilience is achieved by truncating the codeblock at each coding pass (possible with the *RESTART* mode switch [26]), while the lowest protection level is achieved by not truncating a codeblock at all. Experimental results for the best-resilience (fig. 4.4) show that registry-overhead is approximately 47% of codestream size. Also notice that if more than two descriptions are encoded, they do not need to include MQ-encoder registers to provide error-detection capabilities to the central decoder. In fact, the central decoder can deduce which segments are corrupted by using the majority-rule over all registers of available descriptions.

4.2 MDC of WT-Subbands using JPEG-2000

This MD video coder, described in figure 4.1, follows the scheme proposed in [2] and based on [21][25][28], in which the product code and MAP algorithms are replaced by JPEG-2000. The use

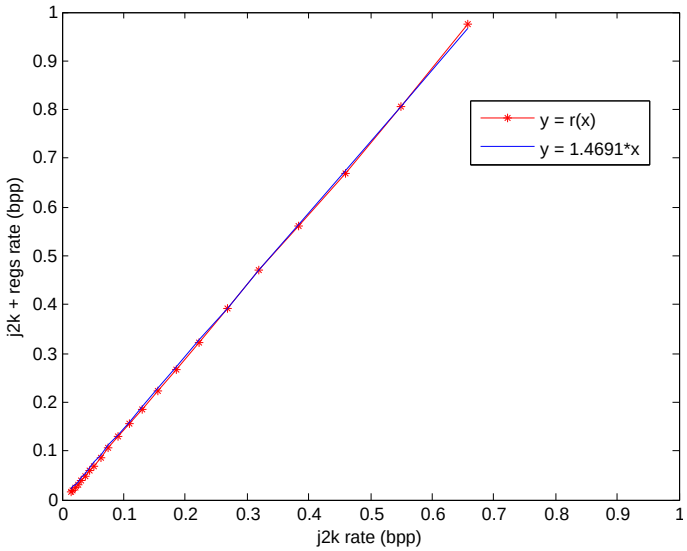


Figure 4.4: Registry-based JPEG-2000 rate in function of standard JPEG-2000 rate: *RESTART* mode switch enabled.

Foreman Sequence ($R_t = 2000$ kbps), AWGN				
$\approx$ BER	R_n	**Side 1**	**Side 2**	**Central**
54×10^{-4}	0.2	22.59	21.86	31.38
24×10^{-4}	0.2	25.44	27.11	34.42
6×10^{-4}	0.2	31.80	31.67	35.77
Foreman Sequence ($R_t = 1500$ kbps, BSC)				
$\approx$ BER	R_n	**Side 1**	**Side 2**	**Central**
5×10^{-4}	0.2	27.52	27.69	33.40
2×10^{-4}	0.8	32.55	31.83	34.11
5×10^{-4}	0.8	28.25	29.63	33.73
Erik Sequence ($R_t = 1300$ kbps, BSC)				
$\approx$ BER	R_n	**Side 1**	**Side 2**	**Central**
4×10^{-4}	0.2	27.80	28.65	30.99
5×10^{-4}	0.2	27.35	28.01	31.00
3×10^{-4}	0.3	28.86	29.17	31.07

Figure 4.5: Side decoders VS central PSNR (decibels)

of our MQ-coder registry-based JPEG 2000 for temporal subband MD coding yields to an efficient central signal reconstruction while maintaining the excellent performance demonstrated in [28].

The video coding scheme starts by performing the temporal motion-compensated wavelet-transform (WT) [28]. The MD bit-allocation (fig. 4.1) efficiently distributes resources to descriptions based on the redundancy parameter [21] and temporal subband rate-allocation is achieved by a rate-distortion algorithm [28] that includes an MD-bit-allocation (fig. 4.1) that efficiently distributes resources to descriptions based on the redundancy parameter (R_n) [21]. The redundancy parameter aims to adapt redundancy to the current channel characteristics, within the flow. Afterwards, they are encoded with our MD JPEG-2000 scheme using the provided target-rate R and the bitstream is sent over a noisy channel. In the decoding process, an inverse temporal motion-compensated WT takes place and temporal subbands are decoded using the method explained in section 4.1.

4.3 Results

Some experiments were performed and compared with [2] that is a highly performing MDC video codec. For that purpose, three (2,0) temporal decomposition levels with 1/4 pixel motion vectors were used. The source sequence "foreman_cif" was encoded using a total target rate $R_t = R_1 + R_2 = 1500$ kbps and "erik_cif" sequence using a total target rate $R_t = R_1 + R_2 = 1300$ kbps, where only the MQ-encoder-saved register overhead is not accounted for. The Binary Symmetric Channel (BSC) and Additive White Gaussian Noise channel (AWGN) were chosen to simulate the noisy transmission channel and were applied to the J2K codestream only, leaving the video stream headers and motion vectors intact. The JPEG 2000 coder was also tuned for error-resilience, using a (irreversible) 9-7 DWT, while EBCOT codeblocks were size 8x8. The JPEG 2000 RESTART mode switch was explicitly enabled to increase the error-resilience to maximum level.

Figure 4.5 shows how the proposed MD Video Coder based on the central reconstruction method explained in section 4.1 (REG-MDC for short) behaves according to different noise levels and description redundancy values (R_n). The summarized results are the mean PSNR value of all the 144 encoded frames. It is clear that the central decoder is capable of successfully extracting non-corrupt information from the two descriptions, providing an admirable central reconstruction. Also notice that the central quality level is much higher than that of the side decoders, confirming that the noise caused by transmission-errors was indeed corrected by our MDC. Also notice that the method is not limited to BSC, it performs well in different types of transmission channels which is the case of AWGN.

Consider the graph in figure 4.8 that compares the quality of the central reconstruction with [2]. The JPEG 2000 central reconstruction method (REG-MDC) clearly outperforms the MAP algorithm at high bit-error rates while still providing similar results for low bit-error rates. In figures 4.8 and 4.5, the reason why the central reconstruction quality of erik is similar for $BER \approx 5 \times 10^{-4}$ and $BER \approx 4 \times 10^{-4}$ is due to the fact that the global encoding rate $R_t = 1300kbps$ is relatively low for this scheme, which causes some temporal subbands (chrominance only) encoded by JPEG 2000 to be too aberrant. In this case, the video coder does not encode these subbands at all, so the REG-MDC method is not used for them during the central decoding. A visual example of MD-decoding using BSC and AWGN with low and high bit-error rates can be seen in figures 4.6 and 4.7, respectively.

As previously stated, no spatial redundancy-allocation is made (only temporal) and the side decoders were only modified to output description-information to the central decoder, so their behavior is still similar to standard JPEG 2000. Since maintaining JPEG 2000 compatibility is a objective, there was no interest in improving the side decoders; therefore a direct comparison of side reconstruction with [2] is not presented. The side reconstruction quality shown in [2] outperforms the values presented in figure 4.5 since they perform spatial redundancy-allocation.

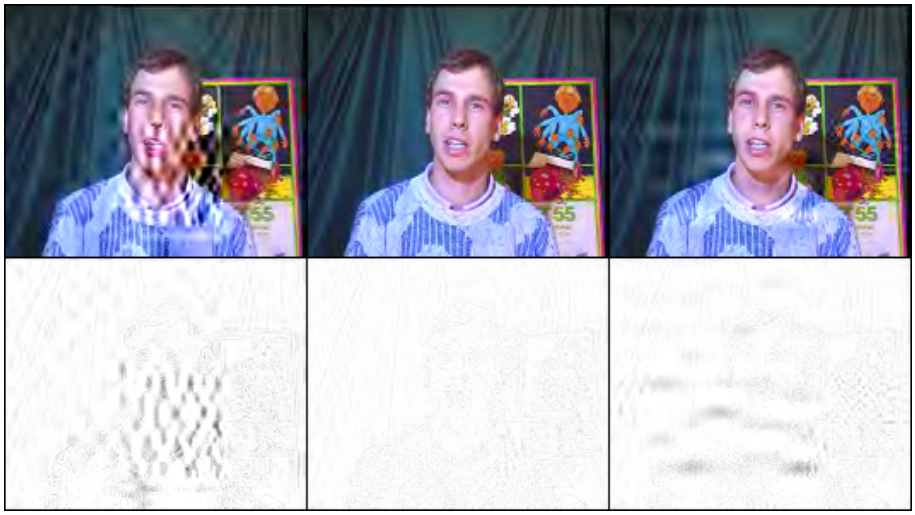


Figure 4.6: Side VS Central reconstruction of frame 113 ("erik_cif" sequence, $R_n = 0.2$, $R_t = 1300$ kbps, $BER \approx 5 \times 10^{-4}$ on a BSC channel). The first line shows the Side 1, Central and Side 2 decoded descriptions and the second line shows their respective visual difference from the original frame.

Figure 4.7: Side VS Central reconstruction of frame 29 ("foreman_cif" sequence, $R_n = 0.2$, $R_t = 2000$ kbps, $BER \approx 54 \times 10^{-4}$ on a AWGN channel). The first line shows the Side 1, Central and Side 2 decoded descriptions and the second line shows their respective visual difference from the original frame.

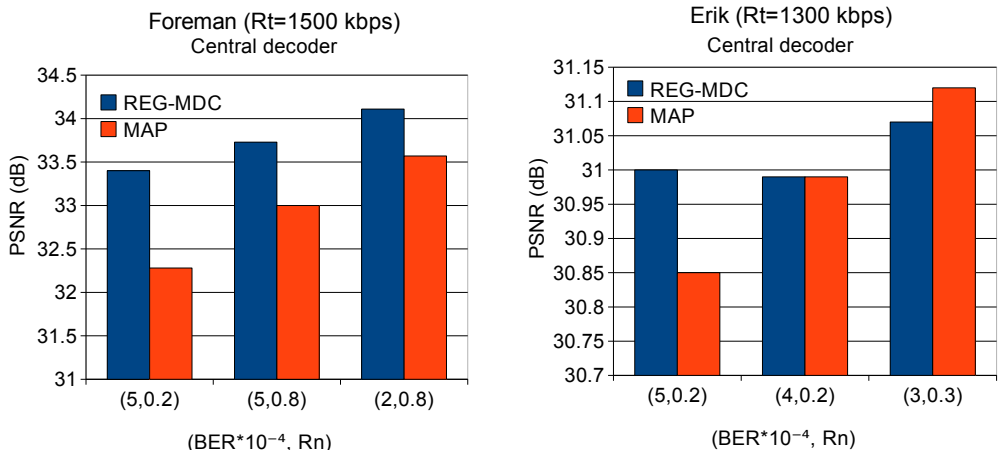


Figure 4.8: Comparison of central reconstruction with [2].

Chapter 5

Multiple-Description-Coding of Motion-Vectors

5.1 MDC using MAP

To study the applications of Multiple-Description-Coding (MDC) using Maximum-A-Posteriori (MAP) for robust decoding of Motion-Vectors (MVs), a computational model was developed. The simulator is written in "C" for greater control on machine code translation and thus faster execution. In figures 5.1 and 5.2 there is a block diagram of the simulation program.

5.1.1 Encoding

Random coefficients $s_0, g_1, g_2 \in \Re$ assembling a Gaussian Probability-Density-Function (PDF) are first generated using the *Box-Muller* algorithm [29]. Then a new set of coefficients for descriptions g_1 and g_2 (fig. 5.1) are generated in such a way that they are correlated with s_0 (with factors c_1 and c_2) and also correlated amongst them. Afterwards, g_1 and g_2 are uniform-scalar quantized, modulated using BPSK and sent through an AWGN channel. The resulting noisy descriptions are r_1 and r_2 which are sent to the decoder.

5.1.1.1 The *Box-Muller* algorithm

Given two uniform-distributed random samples, the *Box-Muller* [29] algorithm is used to generate two Gaussian-distributed samples from them.

First, two uniform-distributed random v_1 and v_2 samples in the interval $[-1, 1]$ are generated. Then $w^2 = (v_1)^2 + (v_2)^2$ is calculated. If $w \notin]0, 1[$, new samples for v_1 and v_2 are generated again. The Gaussian-distributed random samples z_1 and z_2 are generated as follows:

$$z_i = v_i \sqrt{-2 \frac{log(w)}{w}} \tag{5.1}$$

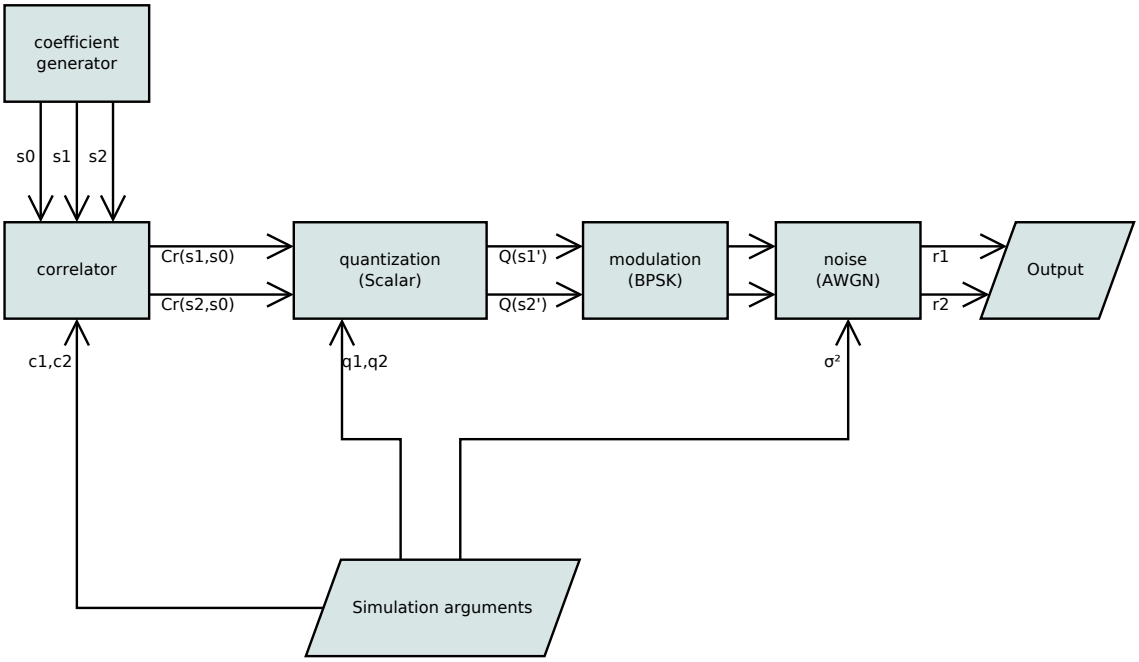


Figure 5.1: MAP MDC: Scheme of simulation program - encoding.

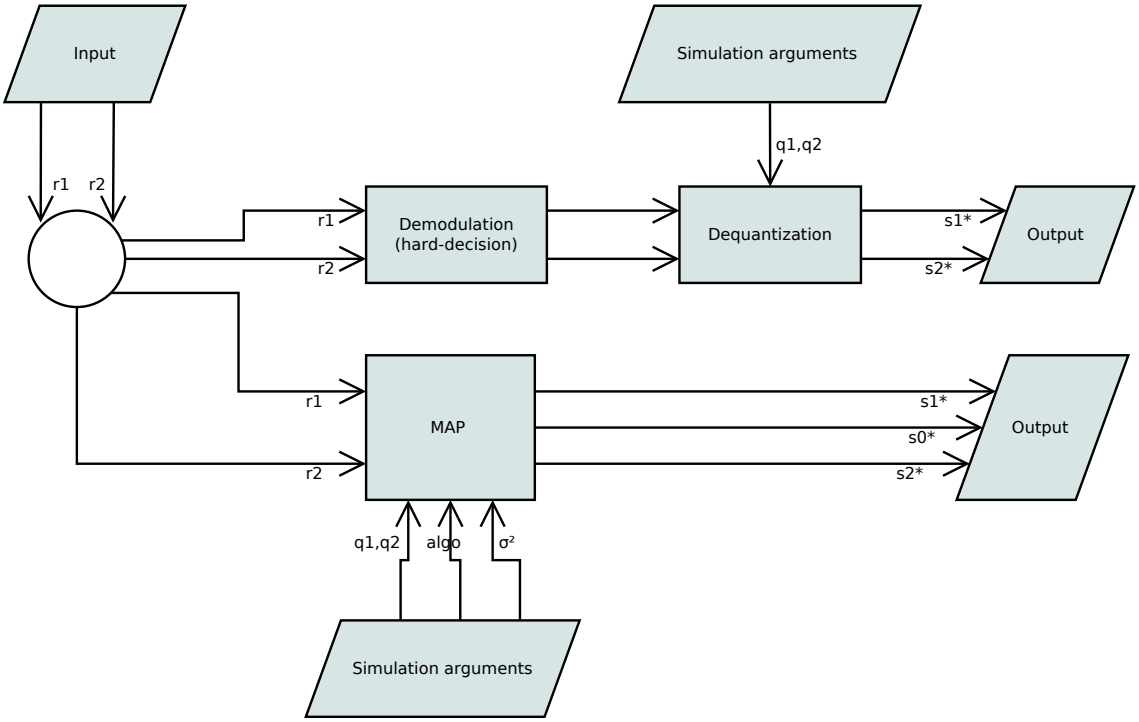


Figure 5.2: MAP MDC: Scheme of simulation program decoding.

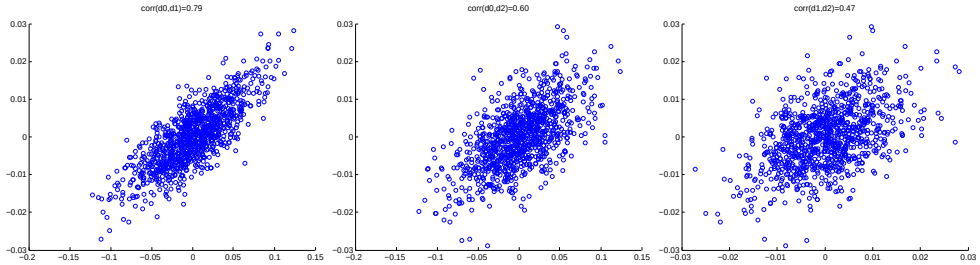


Figure 5.3: From left to right: correlation of g_1 with s_0 ($\rho_1 = 0.8$); correlation of g_2 with s_0 ($\rho_2 = 0.6$) and correlation of g_1 with g_2 ($corr(g_1, g_2) \approx 0.47$). The source s_0 is normal-distributed, zero-centered with $\sigma = 0.2$.

5.1.1.2 Generating correlated variables

One can get y correlated with x_1 with a factor of ρ using the following formula

$$y = \mu + \rho\sigma x_1 + \sqrt{1-\rho^2}\sigma x_2 \tag{5.2}$$

where (μ, σ^2) are the desired normal distribution parameters for y. x_1 and x_2 are both uncorrelated and normal-distributed random variables.

Let s_0 be the source coefficients and g_1, g_2 the side coefficients generated from s_0, they can be generated in a way that they are correlated with the source with factors ρ_1 and ρ_2

$$g_1 = \rho_1 s_0 + \sqrt{1-\rho_1^2} r_1 g_2 = \rho_2 s_0 + \sqrt{1-\rho_2^2} r_2 \tag{5.3}$$

where r_1 and r_2 are normal-distributed random variables. If $\rho_i = 1$ then $\sqrt{1-\rho_i^2} = 0$ which makes the random source have zero influence on g_i and thus $g_1 = s_0$. When $\rho_i = 0$ it implies that $\sqrt{1-\rho_i^2} = 1$, therefore s_0 contributes zero to g_i and thus $g_i = r_i$.

If both $\rho_i \neq 0$ then g_1 and g_2 are correlated with s_0, consequently g_1 and g_2 are also correlated amongst them (fig. 5.3).

5.1.2 Decoding

Noisy descriptions r_1 and r_2 are first hard-demodulated, dequantized and outputted as s_1^* and s_2^*. The result is the unconditional decoding of noisy descriptions, which is later used for comparison. The objective is to reconstruct the original signal s_0 as $\hat{s_0}$ with the lowest possible distortion D_0 using the information provided by the received (noisy) side descriptions r_1 and r_2 (fig. 5.4). Let s_1 and s_2 be the quantized source descriptions before channel transmission, the noisy descriptions are decoded using one of the two MDC-MAP algorithms presented in [3] and explained next.

In the first approach central description $\hat{s_0}$ is obtained by a Maximum-A-Posteriori (MAP) estimate of side descriptions s_j^* ($j = 1, 2$) given the received symbols r_j (fig. 5.5) and a posterior distribution of the source signal, the quantization steps and a channel model. The estimation problem

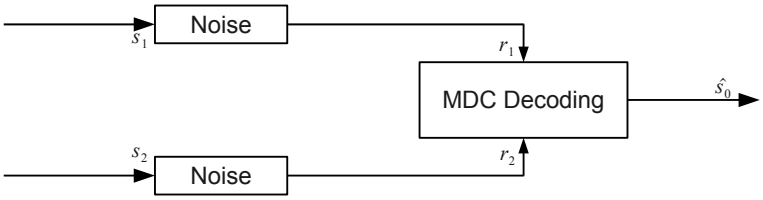


Figure 5.4: General scheme for decoding two noisy descriptions [3].

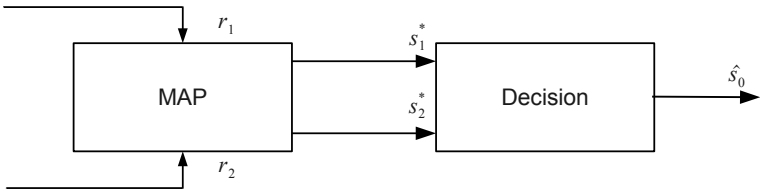


Figure 5.5: MAP first approach [3]: Side descriptions are estimated first.

can be seen as [3]

$$(s_1^*, s_2^*) = arg \max_{s_0} p(s_1, s_2|r_1, r_2) \tag{5.4}$$

which with further simplification [3] using Bayes Rules leads to

$$(s_1^*, s_2^*) = arg \min_{s_1, s_2} -log[p(s_1, s_2)p(r_1|s_1)p(r_2|s_2)] \tag{5.5}$$

where [3]

$$p(s_1, s_2) = \int_{\max(s_1 - \frac{q_1}{2}, s_2 - \frac{q_2}{2})}^{\min(s_1 + \frac{q_1}{2}, s_2 + \frac{q_2}{2})} p_S(x)dx \tag{5.6}$$

The minimization of expression 5.5 leads to the estimation of the optimal side description coefficients s_1^* and s_2^* (fig. 5.5).

The decision to select s_1^* or s_2^* for estimating $\hat{s_0}$ is based on the calculation of the quantization distortions D_1 and D_2 and then choosing the lowest D_i [3]:

$$D_i = \sum_{j \in \beta} p(r_i|s_j) \int_{s_i^* - \frac{q}{2}}^{s_i^* + \frac{q}{2}} (x - s_i^*)^2 p_S(x)dx \tag{5.7}$$

where β is the Quantization codebook, q is the quantization step, $p_S(x)$ is the PDF of the source signal.

To compute equation 5.5 one also needs to calculate $p(r_i|s_i)$ which can be given by the channel model [3]. In the case of a zero-mean White Gaussian channel with noise variance σ^2, one has [3]

$$p(r_i|s_i) = \frac{1}{\sqrt{(2\pi\sigma^2)^M}} exp\left(-\frac{|r_i - u(s_i)|^2}{2\sigma^2}\right) \tag{5.8}$$

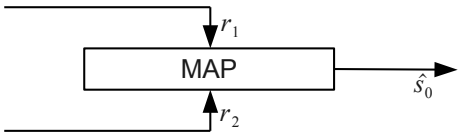


Figure 5.6: MAP second approach [3]: Central coefficients are estimated directly from received noisy descriptions.

where $u(s)$ is a function which represents the source s after quantization, fixed-length M-bit binary indexation, and BPSK signaling.

In the second approach decoding is done by MAP estimation of central description $\hat{s_0}$ (fig. 5.6) directly from received descriptions r_1 and r_2 [3]:

$$\hat{s_0} = arg\max_{s_0} p(s_0|r_1, r_2) \tag{5.9}$$

the development of this expression [3] leads to

$$\hat{s_0} = arg\max_{s_0} \sum_{s_1,s_2} p(r_1|s_1)p(r_2|s_2)p(s_1, s_2|s_0)p(s_0) \tag{5.10}$$

where $p(s_0)$ is the PDF of the source signal and $p(r_i|s_i)$ is the channel model. $p(s_1, s_2|s_0) = 1$ if s_0 belongs to the intersection of the quantization intervals associated to s_1 and s_2, otherwise $p(s_1, s_2|s_0) = 0$ [3].

The maximization of expression 5.10 leads to the estimation of $\hat{s_0}$.

5.1.3 Results

Looking at figure 5.8 the MAP1 $\hat{s_0}$ decision based on estimated descriptions s_j^* is visible and allows the comparison between MAP approaches when using uniform quantization. Indeed, MAP1 always chooses the best available estimated description or, in this case, description s_2^* since it was quantized with smaller quantization steps. The distance along the vertical axis between s_1^* and s_2^* corresponds to the quantization error increase between the two descriptions.

The main difference between MAP1 and MAP2 is that on MAP2 the estimated $\hat{s_0}$ does not need to belong to the quantization-codebook, in fact it can assume any real value inside the subinterval given by the intersection of both description's intervals:

$$\hat{s_0} \in (P_1 \cap P_2), \quad P_1, P_2 \subset \Re \tag{5.11}$$

This can be exploited in such a way that the estimated $\hat{s_0}$ is closer to the original unquantized value s_0 than the quantized values s_1 and s_2.

Experimental results (fig. 5.8) tell us that MAP2 provides as much reconstruction quality as MAP1 if uniform-scalar-quantization is used with a Codebook size $M = 2^m$ symbols (fig. 5.7 left); and the estimated s_0^* is chosen to be the centroid of the subinterval $(P_1 \cap P_2)$. In this way, looking at

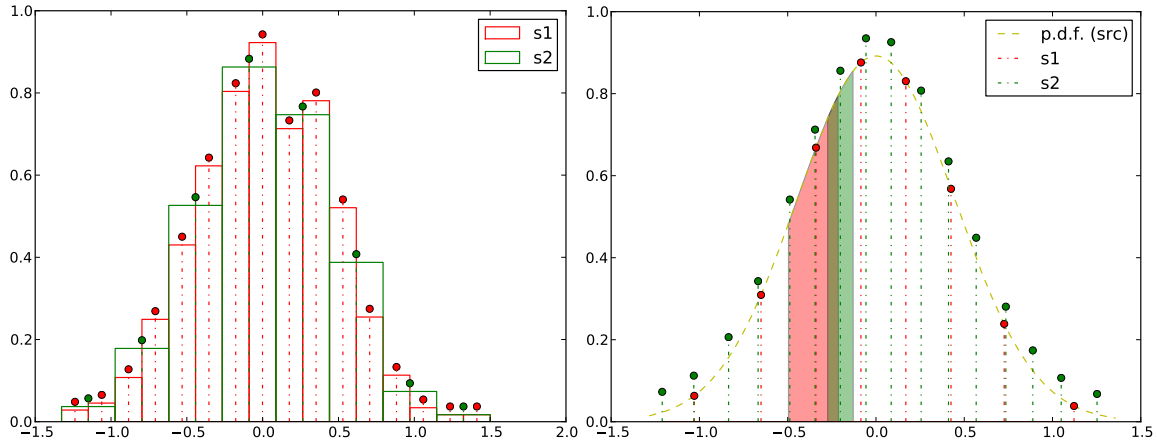


Figure 5.7: PDF, quantization intervals and its representatives. On the left uniform-quantization, on the right quantization using a optimal codebook generated by Loyd's algorithm.

the left of fig. 5.7, it is clear that s_0^* can either assume the value of s_1 or s_2 because they always lay on the interval's boundaries and center. More specifically, the representative of the largest quantization interval will always lay on the boundaries of the subinterval and the representative of the smallest quantization interval will always be the centroid of the subinterval. Therefore, by always selecting $\hat{s_0}$ as the subinterval centroid, the best available description for $\hat{s_0}$ is implicitly chosen; this is the same behavior in MAP1.

However, by modifying the simulation to generate an optimal quantization-codebook through Loyd's algorithm (fig. 5.7 right) and by forcefully modifying the distortion function 5.7 of the MAP1 model to

$$D_i = \sum_{j \in \beta} p(r_i|s_j) \int_{L(s_i^*)}^{U(s_i^*)} (x - s_i^*)^2 p_S(x) dx \tag{5.12}$$

being $U(s)$ and $L(s)$ functions that gives the upper and lower limits of a symbol s based on the quantization codebook β, a special case can happen. This special case results in a subinterval $P_1 \cap P_2$ where both s_1 and s_2 do not belong (fig. 5.7 right). Since MAP1 is constrained to select s_0^* as either s_1^* or s_2^*, by letting s_0^* assume any $\Re$ value within the subinterval $P_1 \cap P_2$ which minimizes the MAP2 criterion (minimization is made through bisection), it was expected that MAP2 would perform a better reconstruction than MAP1. Looking at figure 5.9 it becomes cleat that the former claim does not hold true.

Since both MAP approaches require that the intervals of both descriptions intersect, the correlation factors must necessarily be equal to one $\rho_1 = 1$, $\rho_2 = 1$; due to the way the correlation function (eq. 5.3) maps the values. Hence, before quantization, each description must be a 1:1 copy of the source signal.

5.2 MDC of MVs using MAP

It was previously seen that the quality of central reconstruction provided by MAP ($\overset{*}{s_0}$) depends, apart from some other things, on the quality of prior information of the source s_0 and channel. Because

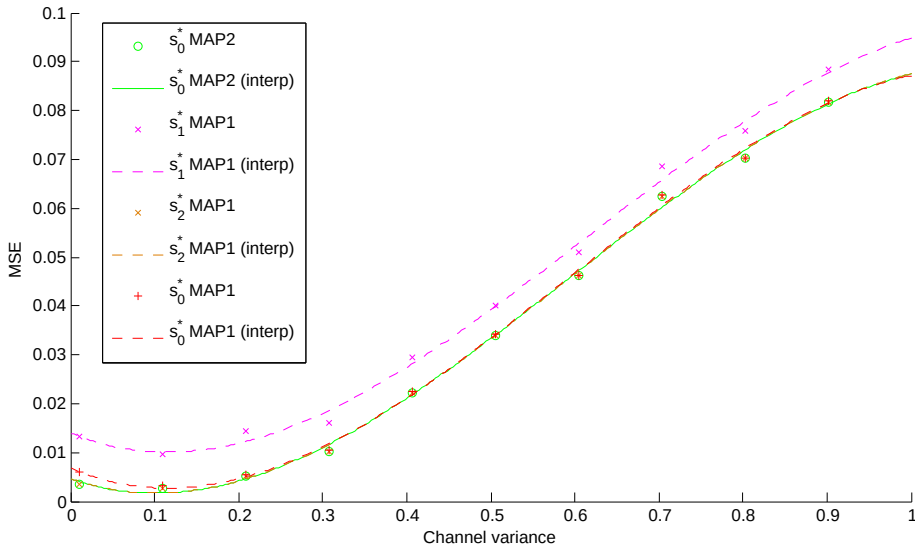


Figure 5.8: Comparison between MAP1 and MAP2: uniform quantization, codebook sizes $M_1 = 2^3$ and $M_2 = 2^4$.

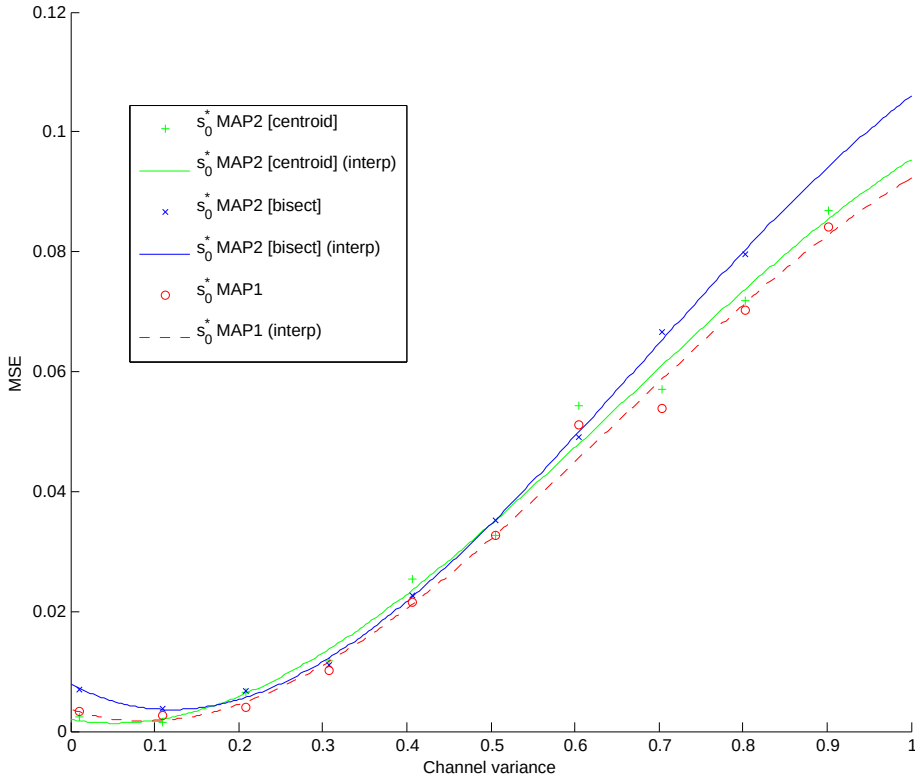


Figure 5.9: Comparison between MAP1 and MAP2: codebook was optimally generated using k-means, codebook sizes $M_1 = 2^3$ and $M_2 = 2^4$.

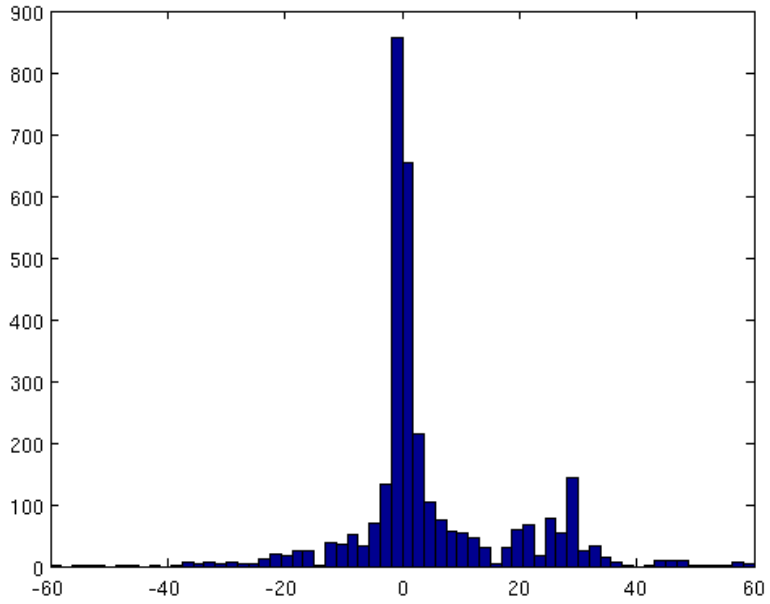


Figure 5.10: Distribution of MVs components from first GOP of the "Edberg" sequence.

BPSK transmission channels subject to Additive-White-Gaussian-Noise are being considered, the channel transition probability density function is Gaussian. Moreover, since in most video sequences the motion happens mostly in a common direction and speed, the distribution of MVs from a Group-of-Pictures (GOP) (fig. 5.10) can be approximated to a Laplacian distribution, being the bell center located in the zone with the mean vector component value. The scale parameter of the Laplace function can also be approximated by the block-matching search window size or maximum motion vector component value. However, in this case which abdicates entropy-coding, it is obvious that reconstruction error is solely introduced by quantization (ignoring channel error). Thereof, it is very important to adapt the quantizer to the source. A solution would be to optimally generate the quantization codebook using Lloyd's algorithm [30]; however the encoder would then need to transmit the entire codebook which has a significant impact on the video rate. Another solution might be the use of the proposal in [31], which is left unstudied in this report due to time constraints. Since vector-coding is restricted to fixed-length coding, for optimality the codebook must be radix-2 of size. With that even number of symbols, it is adequate to use a midrise uniform scalar quantizer. However, it is known that midtread quantizers are most suitable for Gaussian-like distributions, since these exhibit a higher density at the p.d.f. center [32][31]. Considering that the source s_0 p.d.f parameters (mean and scale values) can be transmitted to the decoder with a negligible impact, it is possible to use a simple but better suited midtread scalar quantizer by taking account of the following :

- the quantization deadzone is located in the mean MV component value;
- the MV distribution is not usually symmetric around its mean value (fig. 5.10);
- vector components need to be rounded to existing pixel and sub-pixel space.

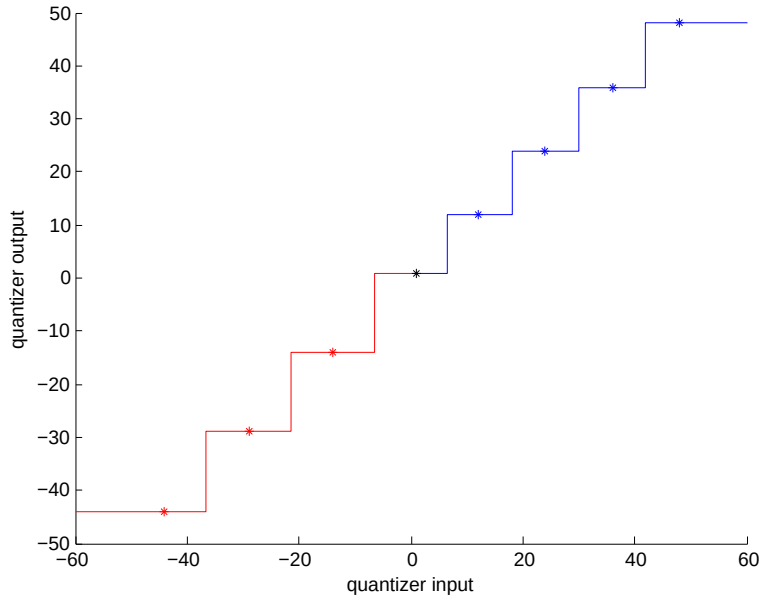


Figure 5.11: Quantizer I/O for MVs components from first GOP of "Edberg" sequence, 1/4 pixel interpolation, 3 bits per vector component.

The quantizer, which needs (in addiction to p.d.f. parameters) the transmission of a single bit to the decoder indicating the Most-Common-Sign (MCS) of MVs components (positive or negative), allocates N quantization representatives in the following way:

- 1 representative for the dead-zone, located in the mean MV component value;

- $\frac{N}{2}$ levels for the most common MV component sign;

- $\frac{N}{2} - 1$ levels for the other sign;

- quantized values are rounded to the next integer towards zero.

For example, consider the MVs components distribution of figure 5.10, where the most common vector-component sign is the positive; in figure 5.11 the corresponding quantizer input/output is visible. There is a total of one quantization level for the dead-zone, three quantization levels for the negative sign and four quantization levels for the positive sign. The quantization overloading towards infinity is purposeful and is caused by the rounding operation.

In short, at decoding the source model, the channel model and its parameters are known a-priori. The parameters of the source model (μ - mean, b - scale) are transmitted. The quantization codebook can be built using transmitted parameters μ, b and MCS. Therefore, the MAP model is complete and can be used to decode the descriptions.

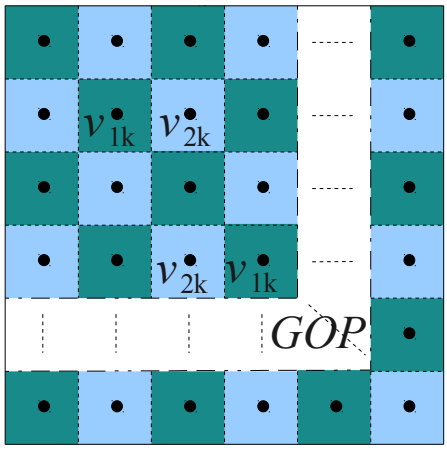


Figure 5.12: Quantization step interleaving for description k

5.3 Bit-allocation for Multiple-Descriptions

As in [21], the problem of efficiently allocating bits to each description, in such a way that central distortion D_0 is minimized, is subject to the description balance constraint

$$R_k = \frac{R_t}{2} \tag{5.13}$$

where R_k is the rate of description k and R_t the total rate.

Considering the following separation of Motion-Vectors (MVs) into two non-intersecting subsets (figure 5.12)

$$v_1 \cup v_2 = v_{gop} \quad v_1 \cap v_2 = \emptyset \tag{5.14}$$

with v_{gop} corresponding to all the MVs in a GOP. The subsets v_1, v_2 are then quantized by their corresponding quantization steps q_1, q_2 which gives the rates R_1, R_2 for each vector group. Since the main constraint is to keep descriptions balanced (eq. 5.13), it is possible to reformulate as

$$R_{1k} + R_{2k} = \frac{R_t}{2} \tag{5.15}$$

Dividing the vectors into two groups now allows the introduction of a new constraint which enforces the similarity between descriptions:

$$R_{11} = R_{22} \tag{5.16}$$

By introducing Lagrange multipliers, the minimization criterion of D_0 subject to the previous constraints is expressed as

$$\begin{array}{c} J_{\lambda_1,\lambda_2,\lambda_3}(R_{11}, R_{12}, R_{21}, R_{22}) = D_0(R_{11}, R_{12}, R_{21}, R_{22}) \\ + \sum_{k=1}^{2} \lambda_k \left[R_{1k} + R_{2k} - \frac{R_t}{2} \right] + \lambda_3 \left(R_{11} - R_{22} \right)^2 \end{array} \tag{5.17}$$

Because J is convex, its global minimum is the point where the slope is zero. More formally, the minimum is the solution of

$$\nabla_{R_{11},R_{12},R_{21}R_{22},\lambda_1,\lambda_2,\lambda_3} J_{\lambda_1,\lambda_2,\lambda_3}(R_{11},R_{12},R_{21},R_{22}) = 0 \tag{5.18}$$

which is given by

$$\begin{cases} \dfrac{\partial J}{\partial R_{11}} = \dfrac{\partial D_0}{\partial R_{11}} + \lambda_1 + 2\lambda_3 (R_{11} - R_{22}) = 0 & i \\ \dfrac{\partial J}{\partial R_{21}} = \dfrac{\partial D_0}{\partial R_{21}} + \lambda_1 = 0 & ii \\ \dfrac{\partial J}{\partial R_{12}} = \dfrac{\partial D_0}{\partial R_{12}} + \lambda_2 = 0 & iii \\ \dfrac{\partial J}{\partial R_{22}} = \dfrac{\partial D_0}{\partial R_{22}} + \lambda_2 + 2\lambda_3 (R_{11} - R_{22}) = 0 & iv \\ \dfrac{\partial J}{\partial \lambda_1} = R_{11} + R_{21} - \dfrac{R_t}{2} = 0 & v \\ \dfrac{\partial J}{\partial \lambda_2} = R_{22} + R_{12} - \dfrac{R_t}{2} = 0 & vi \\ \dfrac{\partial J}{\partial \lambda_3} = (R_{11} - R_{22})^2 = 0 & vii \end{cases} \tag{5.19}$$

Looking at equation system 5.20, it is known that $R_{11} = R_{22}$, which allows further simplification

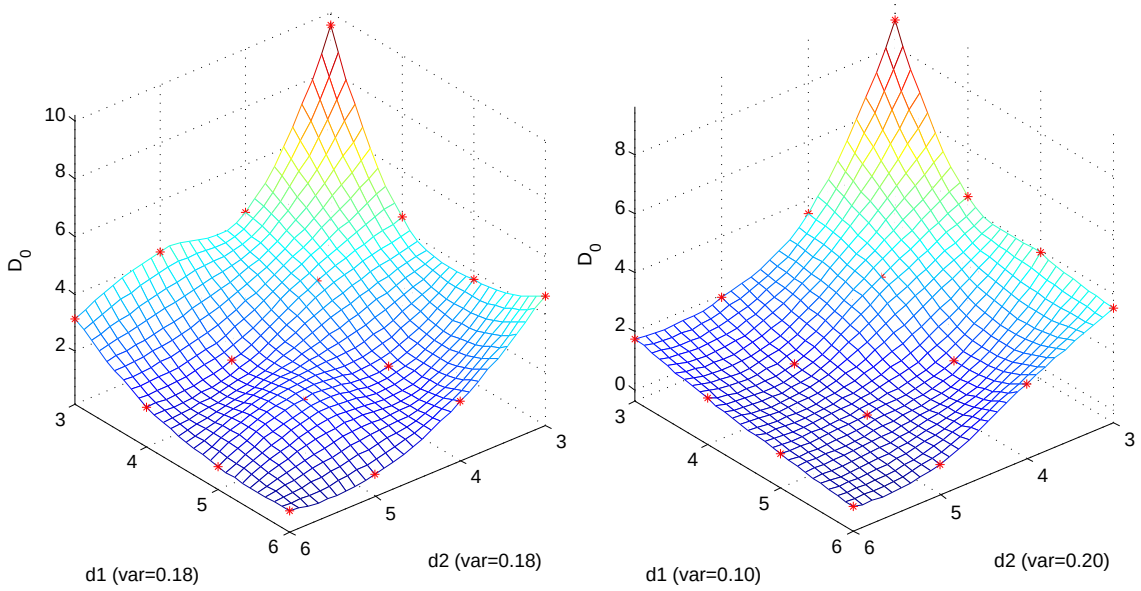
$$\begin{cases} \lambda_1 = -\dfrac{\partial D_0}{\partial R_{11}} \\ \lambda_1 = -\dfrac{\partial D_0}{\partial R_{21}} \\ \lambda_2 = -\dfrac{\partial D_0}{\partial R_{12}} \\ \lambda_2 = -\dfrac{\partial D_0}{\partial R_{22}} \\ R_{21} = \dfrac{R_t}{2} - R_{11} \\ R_{12} = \dfrac{R_t}{2} - R_{22} \\ R_{11} = R_{22} \end{cases} = \begin{cases} \lambda_1 = -\dfrac{\partial D_0}{\partial R_{11}} = -\dfrac{\partial D_0}{\partial R_{21}} \\ \lambda_2 = -\dfrac{\partial D_0}{\partial R_{12}} = -\dfrac{\partial D_0}{\partial R_{22}} \\ R_{21} = \dfrac{R_t}{2} - R_{11} \\ R_{12} = \dfrac{R_t}{2} - R_{11} \end{cases} \tag{5.20}$$

Having found equation 5.20, the search-space for the minimization of criterion 5.17 is now greatly reduced, enabling to minimize the criterion numerically.

5.3.1 Least Squares Minimization

To numerically solve equation 5.20, it is required to estimate $D_0(R_{11}, R_{12}, R_{21}, R_{22})$. Since the two vector groups are independent, it is practical to stochastically estimate the central distortion of a single vector group, as will be seen in the next section. Therefore

$$D_0(R_{11}, R_{12}, R_{21}, R_{22}) = D_{10}(R_{11}, R_{12}) + D_{20}(R_{21}, R_{22}) \tag{5.21}$$



(a) when both channels have same σ

(b) when channels have different σ

Figure 5.13: Estimated R/D surface of a vector group reconstruction at central decoder output.

For calculating the gradient of $D_{g0}(R_{g1}, R_{g2})$, being g the vector group, it is proposed to first evaluate D_{10} and D_{20} at a few bitrate points and then perform their 2D cubic interpolation as visible in figure 5.13. Having an approximation of the rate-distortion surface allows the full-search of the remaining solution-space of 5.20. In short:

- A discrete approximation of $D_0(R_{11}, R_{12}, R_{21}, R_{22})$ is available.

- The achievable rates $R_{11}, R_{12}, R_{21}, R_{22}$ are constrained to a subset of $\Re$ because motion vectors are encoded with a fixed length.

Considering the function $R_{jk} = g(q_{jk})$, which maps quantization steps to the bitrate of motion vectors v_{jk}, it is now necessary to find the point that is closest to the solution of equation 5.20, in function of the quantization steps. For that, equation 5.25 can be rewritten as

$$\begin{cases} \lambda_1 = -\dfrac{\partial D_0}{\partial g(q_{11})} = -\dfrac{\partial D_0}{\partial g(q_{21})} & i \\ \lambda_2 = -\dfrac{\partial D_0}{\partial g(q_{12})} = -\dfrac{\partial D_0}{\partial g(q_{22})} & ii \\ g(q_{21}) = \dfrac{R_t}{2} - g(q_{11}) & iii \\ g(q_{12}) = \dfrac{R_t}{2} - g(q_{11}) & iv \end{cases} \tag{5.22}$$

However, $\nabla_{g(q_{jk})} D_0$ being a discrete model whose domain is the contra-domain of g, equations i

and *ii* in 5.22 cannot always be verified. Consequently, minimization is achieved at the point where

$$-\frac{\partial D_0}{\partial g(q_{11})} \approx -\frac{\partial D_0}{\partial g(q_{21})} \tag{5.23}$$

$$-\frac{\partial D_0}{\partial g(q_{12})} \approx -\frac{\partial D_0}{\partial g(q_{22})} \tag{5.24}$$

The optimal point can then easily be found by means of a least squares minimization

$$\begin{aligned} arg\,min_{q_{11},q_{12},q_{21}\in\Re} &\left(-\frac{\partial D_0}{\partial g(q_{11})}+\frac{\partial D_0}{\partial g(q_{21})}\right)^2+\left(-\frac{\partial D_0}{\partial g(q_{12})}+\frac{\partial D_0}{\partial g(q_{11})}\right)^2 \\ &+\left(g(q_{21})-\frac{R_t}{2}+g(q_{11})\right)^2+\left(g(q_{12})-\frac{R_t}{2}+g(q_{11})\right)^2 \end{aligned} \tag{5.25}$$

5.3.2 Stochastic estimation of D_0

The estimation of central distortion D_0 can be seen as a problem of guessing what the central-decoded vector would be, given that the vector v was encoded and then calculating its associated quantization distortion. Given a vector v at the encoder, the corresponding vector at description k just before transmission is

$$v_k = Q(v, q_k) \quad q_k \in \beta_k \tag{5.26}$$

where Q is the quantizer function, q_k is the quantization step and β_k the quantization codebook.

Given a model of the channel, one can find the maximum likelihood of receiving vector $\overset{*}{r_k}$ given that v_k was transmitted (v_k is the vector to be encoded, it becomes known during bit-allocation):

$$\overset{*}{r_k} = \underset{r_k\in\beta_k}{arg\,max}\; p_k(r_k|v_k) \tag{5.27}$$

Considering a AWGN channel with variance σ^2, the model is

$$p_k(v_k|r_h) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{|v_k - r_h|^2}{2\sigma^2}\right) \tag{5.28}$$

However, for a large number of vectors in a GOP it is more efficient to instead use the Monte Carlo probabilistic approach to sample $\overset{*}{r_k}$ by modulating, adding white Gaussian noise and demodulating v_k.

Having sampled $\overset{*}{r_k}$, the problem now lies in estimating what the central decoded vectors $\overset{*}{v_k}$ would be, if $\overset{*}{r_1}, \overset{*}{r_2}$ were to be received. This estimation is given by the MAP model itself [3]:

$$(\overset{*}{v_1}, \overset{*}{v_2}) = \underset{h_1,h_2}{arg\,min}\; -log\left[p(h_1,h_2)\,p(\overset{*}{r_1},h_1)\,p(\overset{*}{r_2},h_2)\right] \tag{5.29}$$

The distortion of the side decoder k is

$$D_k(v) = (\overset{*}{v_k} - v)^2 \tag{5.30}$$

Since MAP1 will always chose the description symbol that offers the lowest distortion, the central

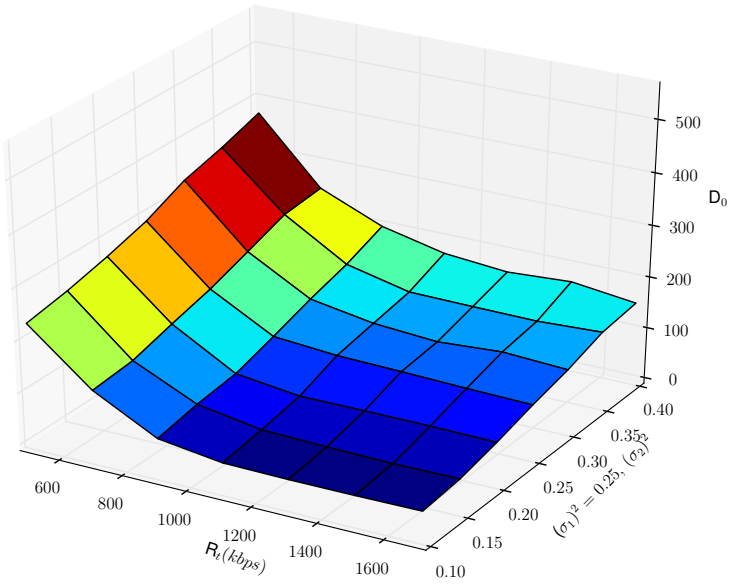


Figure 5.14: R/D curves of MV MDC using the MD Bit-Allocation: Channel 1 variance is fixed at 0.25.

distortion of a vector can then simply be written as

$$\begin{cases} D_1(v) & if\ q_1 \leq q_2 \\ D_2(v) & else \end{cases} \tag{5.31}$$

5.3.3 Results

Figures 5.14 and 5.15 show the output MV distortion at the MD-decoder, in function of the bit-rate and channel variance. Looking with more attention, it is visible that the R/D curves are mostly similar for a channel variance lower than 0.25. However for higher channel variances, the output distortion for the decoder dealing with two channels with different variances is lower. This is indeed the intended behavior for a central MD-decoder and it means that some information is being extracted from the channel with higher capacity (in this case, the channel with lower bit-error-rate). In the MAP context, this behavior is explained by the channel model introduced as prior information.

In figures 5.16 and 5.17, we can compare the reconstruction quality of the full video coder in noisy and noiseless channel environments. For the sake of this test, the total video rate was set to 8Mbps where the MVs were encoded with a variable rate (values along X-axis), while complementary information (subbands, headers) was encoded with the remaining rate. The Motion-Vectors were encoded into two descriptions and transmitted over two different channels, the channels were first setup with noise variances of 0.18 and 0.25 respectively, then setup without any noise at all. The subbands were encoded into four descriptions and always transmitted over noiseless channels. As we can see by looking at the aforementioned figures (5.16 and 5.17), 700kbps is the ideal maximum MV rate for "Edberg" sequence. We can also conclude that the MV Bit-Allocation and MAP decoding algorithms are working as intended, since the central reconstruction is always better than the side

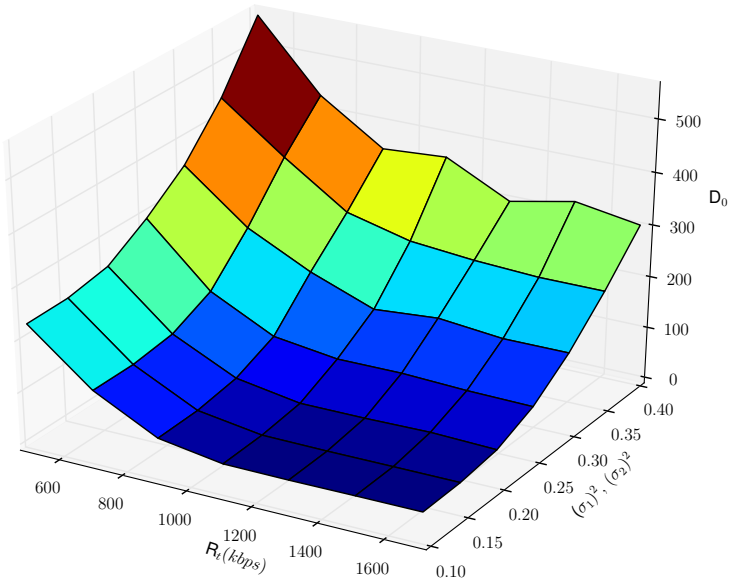


Figure 5.15: R/D curves of MV MDC using the MD Bit-Allocation: Both channels with same variance.

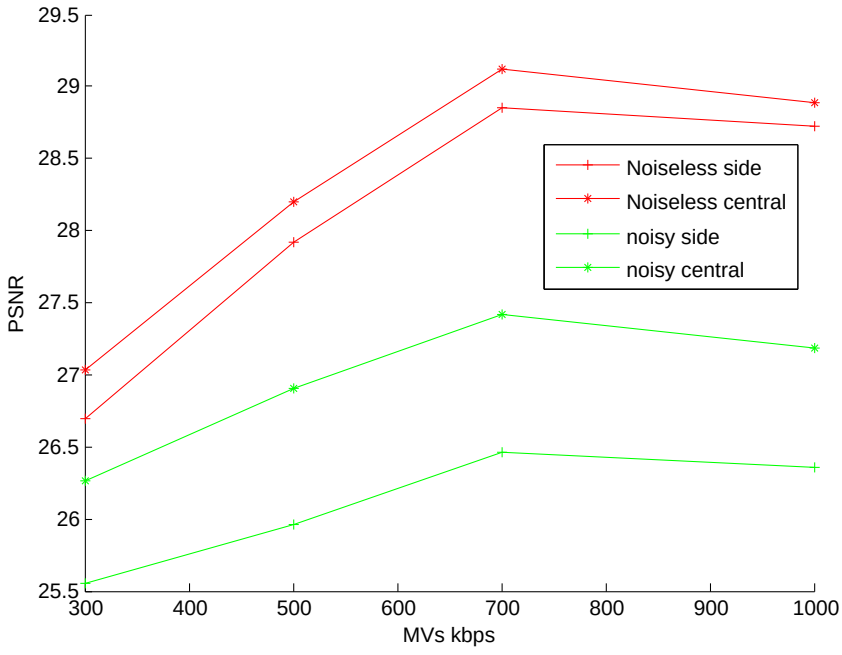


Figure 5.16: Central decoder reconstruction quality for "Edberg" sequence (with variable rate for MVs): comparison of R/D curves for noisy and noiseless channels.

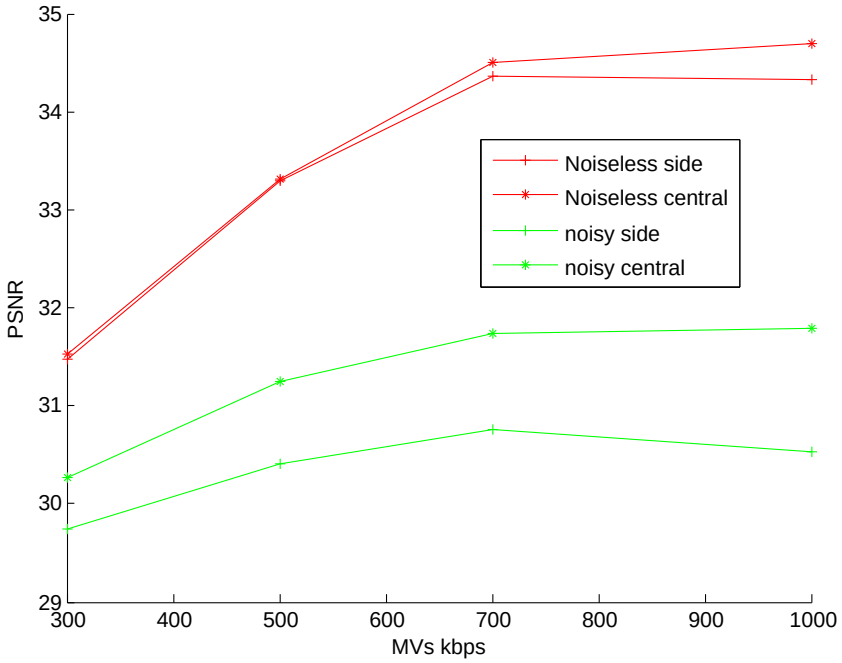


Figure 5.17: Central decoder reconstruction quality for "Foreman" sequence (with variable rate for MVs): comparison of R/D curves for noisy and noiseless channels.

reconstruction, even in noiseless conditions. We can also see that for this case, where channel noise levels are relatively high, the central reconstruction gain is near 1dB.

Chapter 6

Results

For testing the robustness of the entire video coder, the sequence "edberg" was chosen since it is a scene with considerable movement. The robust video coder operates directly over two simulated channels prone to AWGN with relatively high variances 0.18 and 0.25. For comparison, the target bit-rate for the encoded video was chosen to be 500 kb/s.

As visible in table 6, the central reconstruction always provides better results than the best side reconstruction. Average central improvement in this sequence goes up to 1.78 dB and the average reconstruction quality is 18.72 dB, which is rather low. However, notice in the figure 6.2 that the side reconstructions are nearly impossible to visualize: lots of badly compensated blocks and erroneously decoded subbands. The central decoder did, nonetheless, a successful job in recovering most of subband and motion information; being most of distortion caused by source coding. Remains, however, the problem of low central reconstruction quality which can be solved by increasing the video bit-rate. The author acknowledges that the provided quality is akin from current state-of-the-art standards like H.264, however the presented video coder is intended to be deployed in certain environments where those standards will certainly fail to behave: in very noisy wireless links where packet retransmission is not acceptable.

frame	d_1 (dB)	d_0 (dB)	d_2 (dB)	inc. (dB)	best?
1	16.53	17.61	14.38	1.08	YES
2	17.40	18.40	15.40	1.00	YES
3	16.61	18.34	14.91	1.72	YES
4	16.85	18.45	15.15	1.61	YES
5	16.31	18.17	14.71	1.86	YES
6	17.13	18.53	15.07	1.40	YES
7	17.40	19.68	14.61	2.28	YES
8	18.08	20.03	14.55	1.94	YES
9	17.90	20.58	13.76	2.68	YES
10	17.64	19.80	14.70	2.16	YES
11	16.66	19.36	14.32	2.71	YES
12	17.48	19.12	14.79	1.65	YES
13	16.92	18.76	13.93	1.84	YES
14	17.05	18.42	14.13	1.37	YES
15	15.74	17.28	13.08	1.54	YES
16	15.38	17.02	12.91	1.64	YES
avg	16.94	18.72	14.40	1.78	YES

Figure 6.1: Robust video coder: comparison (PSNR) between side and central reconstruction: quality of each frame at each description, central description quality increase relative to side descriptions.

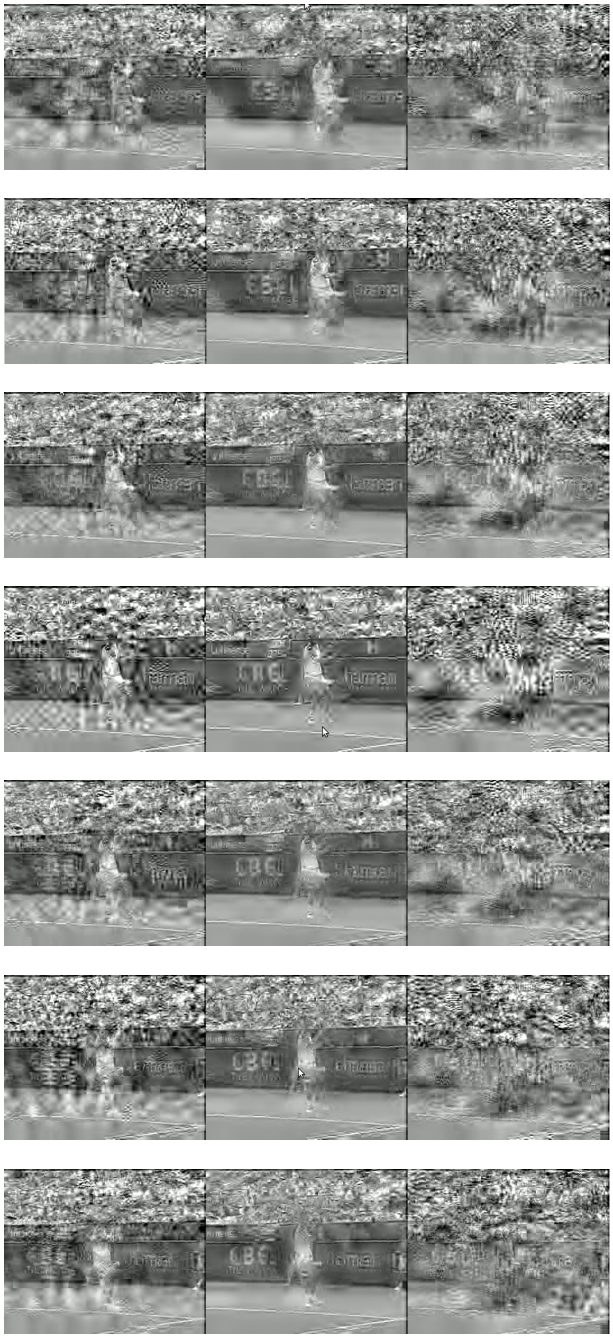


Figure 6.2: Edberg frames 6-12, from top to bottom: side and central descriptions.

Chapter 7

Conclusion

In summary, the combination of two Multiple-Description-Coding algorithms that assure the robust transmission and decoding of information, the efficient resource-allocation between multiple-descriptions based on channel characteristics, the required and highly-efficient Motion-Compensated-Wavelet-Transform and the use of a global resource-allocation between motion and subbands information leads to a highly robust video coding solution with lots of potential for streaming applications over unstable wireless networks, such as 3G broadband. As seen in the presented results, the video coder can be successfully used to transmit motion-compensated video through noisy wireless links. Unlike most of other solutions in the literature, this robust video coder does not target packet erasure channels. On the contrary, because it operates directly in the physical and data-link layers of the OSI model, it can detect and correct transmission bit-errors on the fly without ever requiring data-retransmission. Moreover, direct comparison with high performance video coders like MPEG-4 AVC is not possible since they fail to behave without the use of packet-retransmission.

There are still some important details that need to be addressed, like the robust transmission of bit-stream headers. The presented video coder framework already protects most of information present in a video stream: the spatial and temporal information. For future work, the author proposes exploring FECs as a potential solution to protect header data.

References

[1] Angelo Arrifano, Manuela Pereira, Marc Antonini, and Mário Freire. Multiple description video coding based on JPEG 2000 MQ-coder registers. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010.

[2] Marie Andrée Agostini and Marc Antonini. Multiple description video decoding using MAP. *IEEE International Conference on Image Processing (ICIP)*, October 2008.

[3] Marie A. Agostini. *Nouvelles Approches pour la compression de Videos Haute Definition - Application au Codage par Descriptions Multiples.* PhD thesis, Université de Nice - Sophia Antipolis, July 2009.

[4] V. K. Goyal and J. Kovacevic. Generalized multiple description coding with correlating transforms. *IEEE Trans. on Information Theory*, 2001.

[5] George C. Clark and J. Bibb Cain. *Error-Correction Coding for Digital Communications.* Plenum Press, 1981.

[6] Angelo Arrifano, Manuela Pereira, Marc Antonini, and Mário Freire. JPEG 2000 MQ-coder registry based error detection for lossy transmission channels. In *SPIE Proc. Visual Communications and Image Processing (VCIP)*, pages 1–8, 2009.

[7] Marie Andrée Agostini, Marc Antonini, and Michel Barlaud. Model-based bit allocation between wavelet subbands and motion information in MCWT video coders. *European Signal Processing Conference (EUSIPCO)*, September 2006.

[8] Ya-qin Zhang John Villasenor and Jiangtao Wen. Robust video coding algorithms and systems. *Proceedings of the IEEE*, 87(10):1724–1733, October 1999.

[9] Byeungwoo Jeon Jungyoup Yang. Selective multiple description coding of motion vector with H.264/AVC data partitioning. In *Picture Coding Symposium (PCS)*, November 2007.

[10] Xiuzi Ye Lin Liu, Sanyuan Zhang and Yin Zhang. Error resilience schemes of h.264/avc for 3g conversational video services. In *Fifth International Conference on Computer and Information Technology (CIT'05)*, pages 657 – 661, 2005.

[11] H. Witsenhausen. On source networks with minimal breakdown degradation. *Bell System Technical Journal*, 59:1083 – 1087, Jul 1980.

[12] L. Ozarow. On a source-coding problem with two channels and three receivers. *Bell System Technical Journal*, 59:1909 – 1921, Dec 1980.

[13] A. Wyner J. Wolf and J. Ziv. Source coding for multiple descriptions. *Bell System Technical Journal*, 59:1417 – 1427, Oct 1980.

[14] A. El Gamal and T. M. Cover. Achievable rates for multiple description. *IEEE Trans. Information Theory*, 28(6):851 – 857, Nov 1982.

[15] N. Jayant. Subsampling of a dpcm speech channel to provide two self-contained half-rate channels. *Bell Systems Technical Journal*, 69(4):501 – 509, Apr 1981.

[16] L. Chang-kuan L. Meng-ting and H.H. Chen. Multiple description coding with spatial-temporal hybrid interpolation in peer-to-peer networks. *Journal of Zheijiang University-Science*, 7:894 – 899, 2006.

[17] Manuela Pereira Marisa Quaresma, Angelo M. Arrifano, Mario M. Freire, and Marc Antonini. Robust peer-to-peer video streaming based on multiple description coding. In *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pages 1–6, 2010.

[18] Vaishampayan V.A. Design of multiple description scalar quantizers. *IEEE Transactions on Information Theory*, 39(3):821 – 834, May 1993.

[19] M. Orchard Y. Wang and A. Reibman. Multiple description image coding for noisy channels by paring transform coefficients. *IEEE First Workshop on Multimedia Signal Processing*, pages 419 – 424, June 1997.

[20] R. Puri and K. Ramchandran. Multiple description source coding using forward error correction codes. *Proc. 33rd Asilomar Conf. on Signals, Systems and Computers*, 1:342 – 346, October 1999.

[21] Manuela Pereira. *Multiple Description Image and Video Coding for Noisy Channels*. PhD thesis, Université de Nice - Sophia Antipolis, Jun 2004.

[22] Tammam Tillo, Marco Grangetto, and Gabriella Olmo. Multiple description image coding based on lagrangian rate allocation. *IEEE Transactions on Image Processing*, pages 673 – 683, March 2007.

[23] Tammam Tillo and Gabriella Olmo. Improving the performance of multiple description coding based on scalar quantization. *IEEE Signal Processing Letters*, pages 329–332, 2008.

[24] Jing Wang and Jie Liang. Filter banks for prediction-compensated multiple description coding. In *Data Compression Conference*, pages 392 – 401, Snowbird, Utah, March 2008.

[25] Manuela Pereira, Marc Antonini, and Michel Barlaud. Multiple description image and video coding for wireless channels. *EURASIP Signal Processing: Image Communication*, 18(10):925–945, November 2003.

[26] Frederic Dufaux and Didier Nicholson. JPWL: JPEG 2000 for wireless applications. In *SPIE Proc. Applications of Digital Image Processing XXVII*, 2004.

[27] ISO/IEC. JPEG 2000 image coding system, Mar. 2000.

[28] Marie Andrée Agostini and Marc Antonini. Motion-adapted weighted lifting scheme for MCWT video coders. *Picture Coding Symposium (PCS)*, November 2007.

[29] G. E. P. Box and Mervin E. Muller. A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 29(2):610–611, 1958.

[30] S. P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28:129–137, 1982.

[31] Gary J. Sullivan. Efficient scalar quantization of exponential and laplacian random variables. *IEEE Transactions on Information Theory*, 42(5):1365–1374, 1996.

[32] David L. Neuhoff Robert M. Gray. Quantization. *IEEE Transactions on Information Theory*, 44(6):2325–2383, 1998.