An Improved MPEG-4 Coder Using Lagrangian Coder Control

Heiko Schwarz and Thomas Wiegand
Heinrich Hertz Institute
[schwarz,wiegand]@hhi.de

Summary

This document describes an MPEG-4 coder whose operational mode is rate-distortion optimized using Lagrangian techniques. A similar method has been proposed to ITU-T VCEG by the second author in Q15-D-13[1] for the application to the test model of the ITU-T Recommendation H.263, version 2. The proposal in Q15-D-13 led to the creation of a new encoder recommendation (TMN-10) and is up to now the core of the high-complexity mode of the latest TMN. It should also be noted that the TML, the test model of H.26L, the new project of ITU-T VCEG, basically follows this approach.

The coder generates an MPEG-4 bit-stream compliant with ISO/IEC JTC 1/SC 29 14496-2 (MPEG-4 Visual, version 2). Comparisons are made against the MoMuSys coder according to the VM-17[2] description. The simulations are run for the set of test sequences and conditions as specified in the coding efficiency CfP[3]. All bit-streams have been successfully decoded and results have been derived using the MoMuSys advanced simple profile decoder and an accompanying excel document contains those results.

We have observed that for all sequences tested, the improved encoding strategy provides PSNR gains of between 0.5 and 1.5 dB or, correspondingly, bit-rate savings of about 10-20% at medium bit rates comparing against the MoMySys encoding strategy. When coding at low bit-rates, PSNR gains up to 3 dB and bit-rate savings of up to 40% have been obtained.

1. Motion Estimation and Mode Selection

The problem of optimum bit allocation to the motion vectors and the residual coding in any hybrid video coder is a non-separable problem requiring a high amount of computation. To circumvent this joint optimization, we split the problem of macroblock encoding into two parts: motion estimation and mode decision. For that, motion estimation for the various modes (*INTER*, *INTER-4V*, etc.) is conducted first, and then given these motion vectors, the overall rate-distortion costs for all macroblock modes are computed for rate-constrained mode decision. In the remainder of this section, our approaches to motion estimation and mode decision are described. In the next section, the complete algorithm is given.

[1] ITU-T/SG 16/Q15-D-13 can be obtained via anonymous ftp to standard.pictel.com/video-site/9804_Tam/q15d13.doc.
[2] ISO/IEC JTC 1/SC 29/WG 11, "MPEG-4 Video VM 17.0", Doc. N3515, Beijing, China, July 2000
[3] ISO/IEC JTC 1/SC 29/WG 11, “Call For Proposals On New Tools For Video Compression Technology”, Doc. N4065**,** March 2001, Singapore

1.1 Rate-Constrained Motion Estimation

For each block or macroblock the motion vector is determined by full search on integer-pixel positions followed by half-pixel and possibly quarter-pixel refinement. The integer-pixel search is conducted over the range [-32…32]x[-32…32] pixels relative to the position of the block or macroblock in the current frame. The search is conducted given the predictor of the block or macroblock motion vector.

We view motion-compensated prediction as a source coding problem with a fidelity criterion. For bit-allocation, we use a Lagrangian formulation wherein distortion is weighted against rate using a Lagrange multiplier. More precisely, our integer-pixel motion search as well as our sub-pixel refinement returns the motion vector that minimizes

$$J(\mathbf{m} \mid l_{MOTION}) = SAD(s,\ c(\mathbf{m})) + l_{MOTION} \cdot R(\mathbf{m} - \mathbf{p})$$

with $\mathbf{m} = (m_x, m_y)^T$ being the motion vector, $\mathbf{p} = (p_x, p_y)^T$ being the prediction for the motion vector, and l_{MOTION} being the Lagrange multiplier. The rate term $R(\mathbf{m}-\mathbf{p})$ represents the motion information only and is computed by a table-lookup. The SAD is computed as

$$SAD(s,\ c(\mathbf{m})) = \sum_{x=1,\ y=1}^{B,\ B} \left| s[x,y] - c[x-m_x,\ y-m_y] \right|, \quad B = 16 \text{ or } 8\,.$$

with s being the original video signal and c being the coded video signal. The choice of l_{MOTION} has a rather small impact on the result of the 16x16 block motion estimation. But the search result for 8x8 blocks is strongly affected by l_{MOTION}. In our coder, we choose

$$l_{MOTION} = 0.92 \cdot QP\,,$$

where QP is the macroblock quantization parameter.

1.2 Rate-Constrained Mode Decision

In the proposed coder, the current macroblock mode is chosen given the mode decisions made for the past macroblocks. Rate-constrained mode decision refers to the minimization of the following Lagrangian functional

$$J(s,c,MODE \mid QP, l_{MODE}) = SSD(s,c,MODE \mid QP) + l_{MODE} \cdot R(s,c,MODE \mid QP)$$

where QP is the macroblock quantizer, l_{MODE} is the Lagrange multiplier for mode decision, and $MODE$ indicates a mode chosen from the set of potential prediction modes for progressive video material:

P-VOP: $MODE \in \{INTRA, SKIP, INTER, INTER\text{-}4V\}$,

S-VOP: $MODE \in \{INTRA, SKIP, INTER, INTER\text{-}4V, SPRITE\}$,

B-VOP: $MODE \in \{DIRECT, SKIP, FWDFRM, BAKFRM, AVEFRM\}$.

Note that the $SKIP$ mode refers to the macroblock mode where the COD bit is set to "1" (P-,S-VOP) or the MODB is set to "0" (B-VOP). SSD is the sum of the squared differences between the original block s and its reconstruction c being given as

$$SSD(s, c, MODE \mid QP) = \sum_{x=1, y=1}^{16,16} \left(s[x, y] - c[x, y, MODE \mid QP] \right)^2 ,$$

and $R(s, c, MODE \mid QP)$ is the number of bits associated with choosing $MODE$ and QP including the bits for the macroblock header, the motion, and all six DCT blocks. $c[x, y, MODE \mid QP]$ represents the reconstructed luminance values corresponding to $s[x, y]$. We choose

$$l_{MODE} = 0.85 \cdot QP^2 ,$$

where QP is the macroblock quantization parameter.

2. The Algorithm for Rate-Constrained Encoding

The procedure to encode one macroblock s in a P-, S- or B-VOP in our video codec is summarized as follows.

1. Given the last decoded frame, l_{MODE}, l_{MOTION}, and the macroblock quantizer QP of the VOP of the same type (P, S or B)

2. Perform local motion estimation by minimizing

$$J(\mathbf{m} \mid l_{MOTION}) = SAD(s,\ c(\mathbf{m})) + l_{MOTION} \cdot R(\mathbf{m} - \mathbf{p})$$

   for each motion vector of a possible macroblock mode.

3. Choose the macroblock prediction mode by minimizing

$$J(s, c, MODE \mid QP, l_{MODE}) = SSD(s, c, MODE \mid QP) + l_{MODE} \cdot R(s, c, MODE \mid QP),$$

   given QP and l_{MODE} when varying $MODE$. $MODE$ indicates a mode out of the set of potential macroblock modes for progressive video material. This set depends on the VOP type:

P-VOP: $MODE \in \{INTRA, SKIP, INTER, INTER - 4V\}$,

S-VOP: $MODE \in \{INTRA, SKIP, INTER, INTER - 4V, SPRITE\}$,

B-VOP: $MODE \in \{DIRECT, SKIP, FWDFRM, BAKFRM, AVEFRM\}$.

The computation of $J(s,c,INTRA|QP, l_{MODE})$, $J(s,c,SKIP|QP, l_{MODE})$, and $J(s,c,SPRITE|QP, l_{MODE})$ is simple. The cost for the other prediction modes are computed using the motion vectors, which have been estimated using the rate-constrained motion estimation in step 2.

3. Remaining Coder Parts

All remaining coder parts are operated as described in the MPEG-4 Video Verification Model, version 17.

4. Experimental Results

For illustration purpose of effectiveness of the proposed encoding strategy in comparison to VM 17, parameterized rate-distortion curves are plotted. These curves show PSNR of the luminance component versus bit-rate measured of the complete bit-stream. PSNR is measured as the arithmetic mean of the PSNR values for each frame. The bit-rate is averaged over the complete sequence. The plots have 0.5 dB grid lines on the PSNR axis.

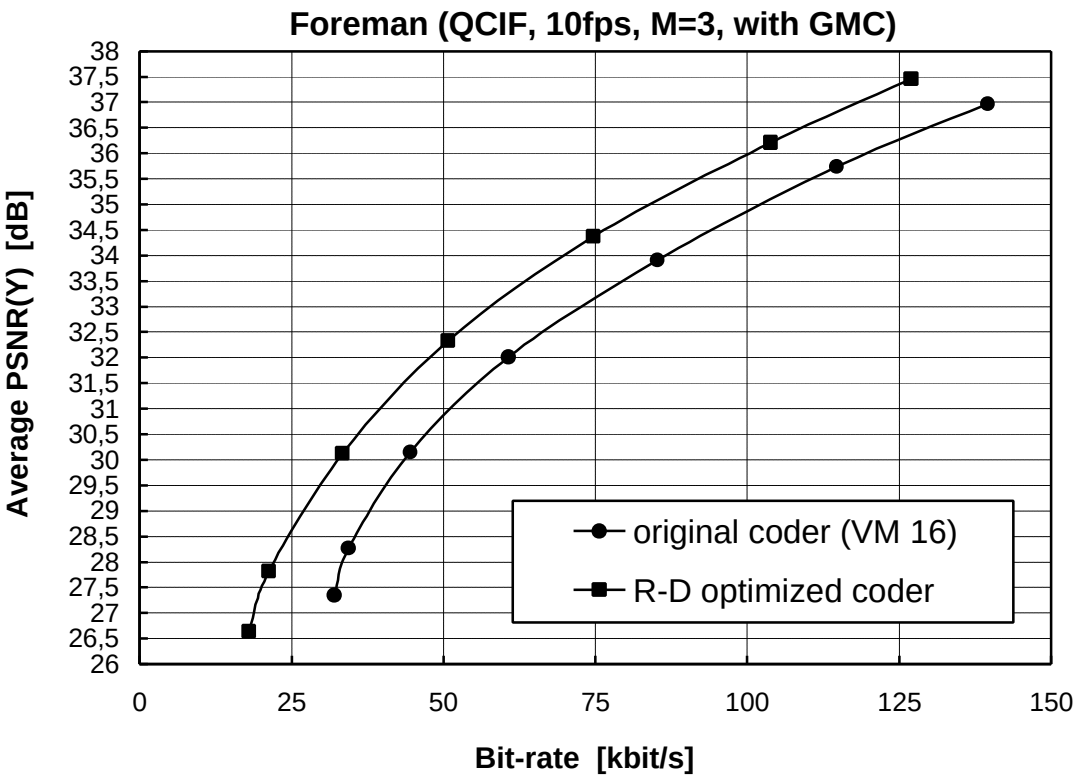
For the following sequences specified in the coding efficiency CfP rate-distortion curves are depicted:

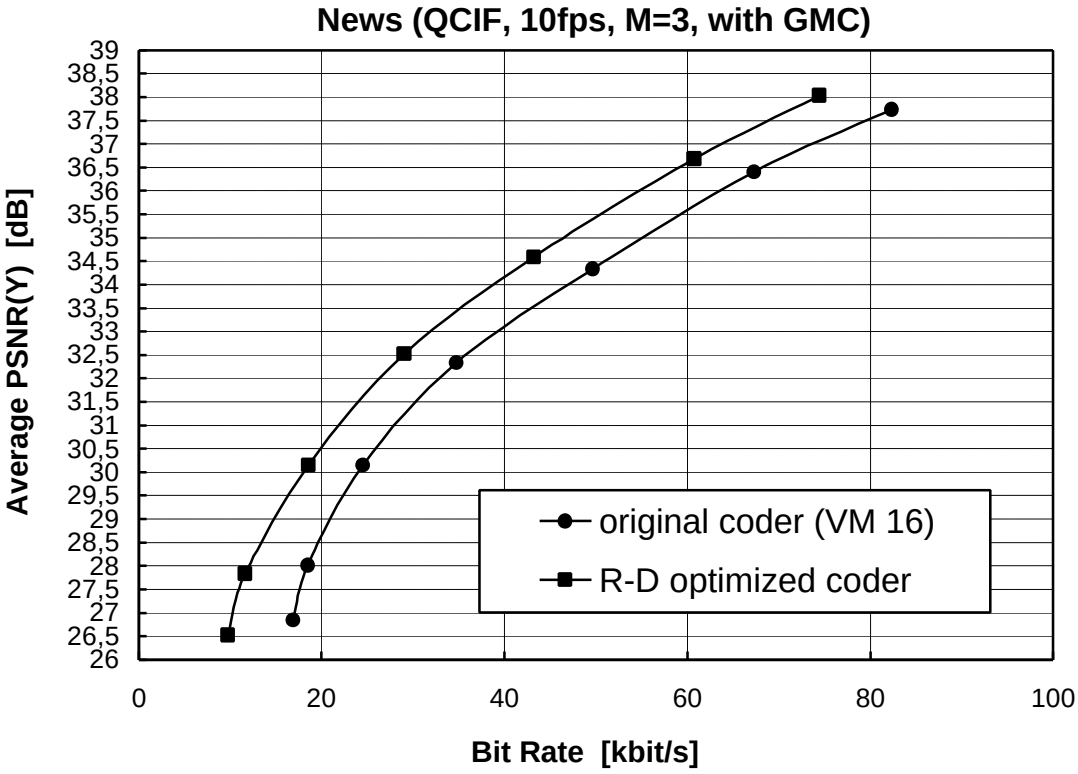
Name	Format	Frame Rate	# of frames
Foreman	QCIF	10 fps	100
News	QCIF	10 fps	100
Tempete	QCIF	10 fps	87
Bus	CIF	15 fps	75
Flowers & Garden	CIF	15 fps	125
News	CIF	15 fps	150

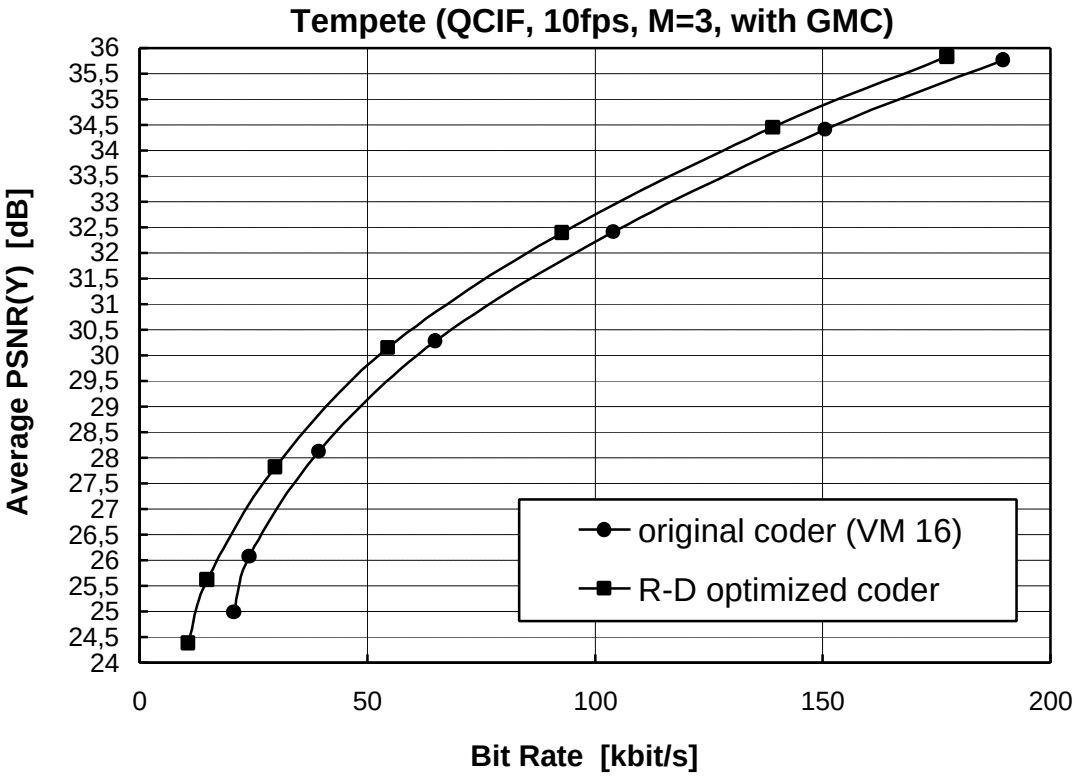
The rate-distortion curves are generated by varying the value of the macroblock quantizer QP that is fixed for a sequence. The parameters given below have been used for the original MoMuSys coder as well as for the rate-distortion optimized coder. At the end of the sequences the distance between two consecutive P- or S-VOPs is adapted so that the last frame is always coded as P- or S-VOP.

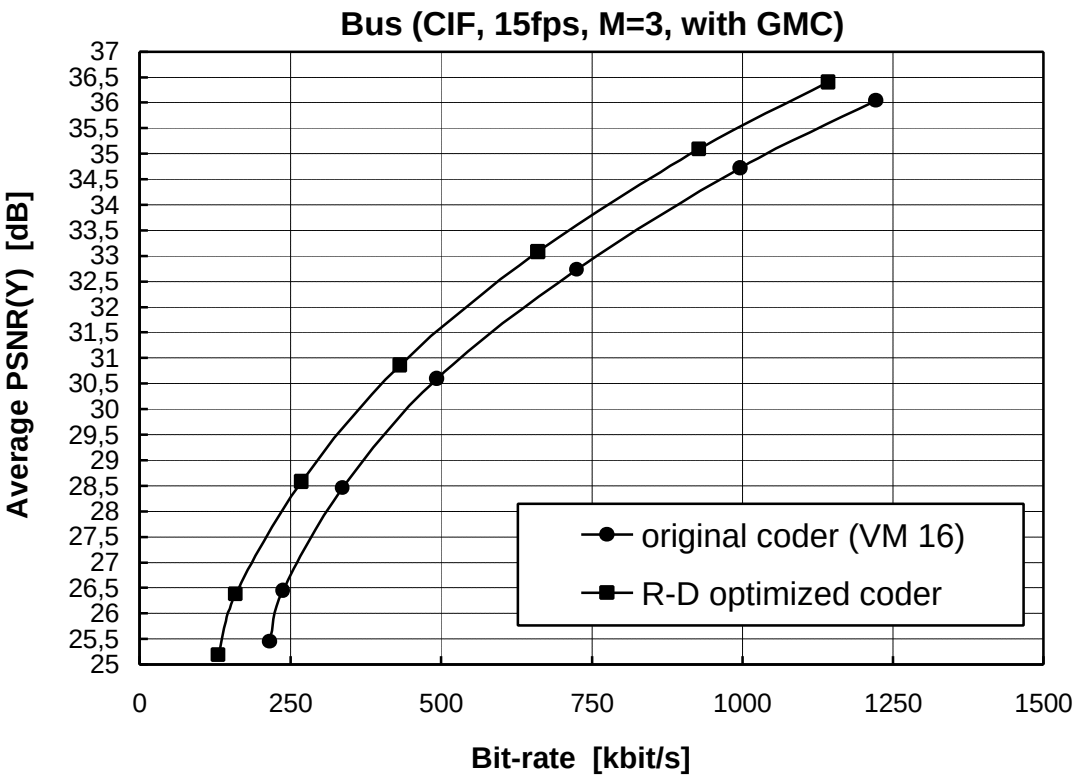
QP_I (I-VOP) and $QP_{P,S}$ (P-, S-VOP)	31, 25, 15, 10, 7, 5,4
QP_B (B-VOP)	$1.2 \cdot QP_{P,S}$
Quantisation matrix	MPEG-2 quantization matrix

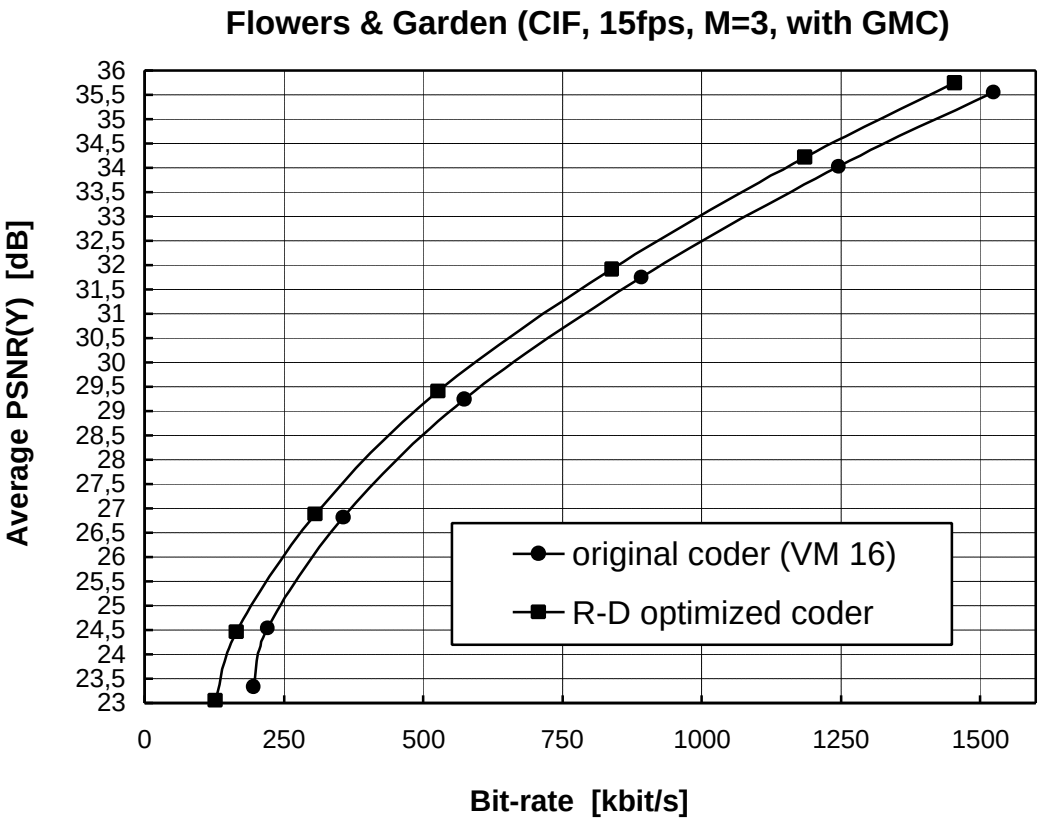
Quarter-pixel MC	enabled
M (distance P-P or S-S)	3 (i.e., 2 B-frames inserted)
GMC	enabled
Search range for LMC	[-32…32]x[-32…32]
Post Filter	Disabled for SNR measurement Enabled for subjective tests

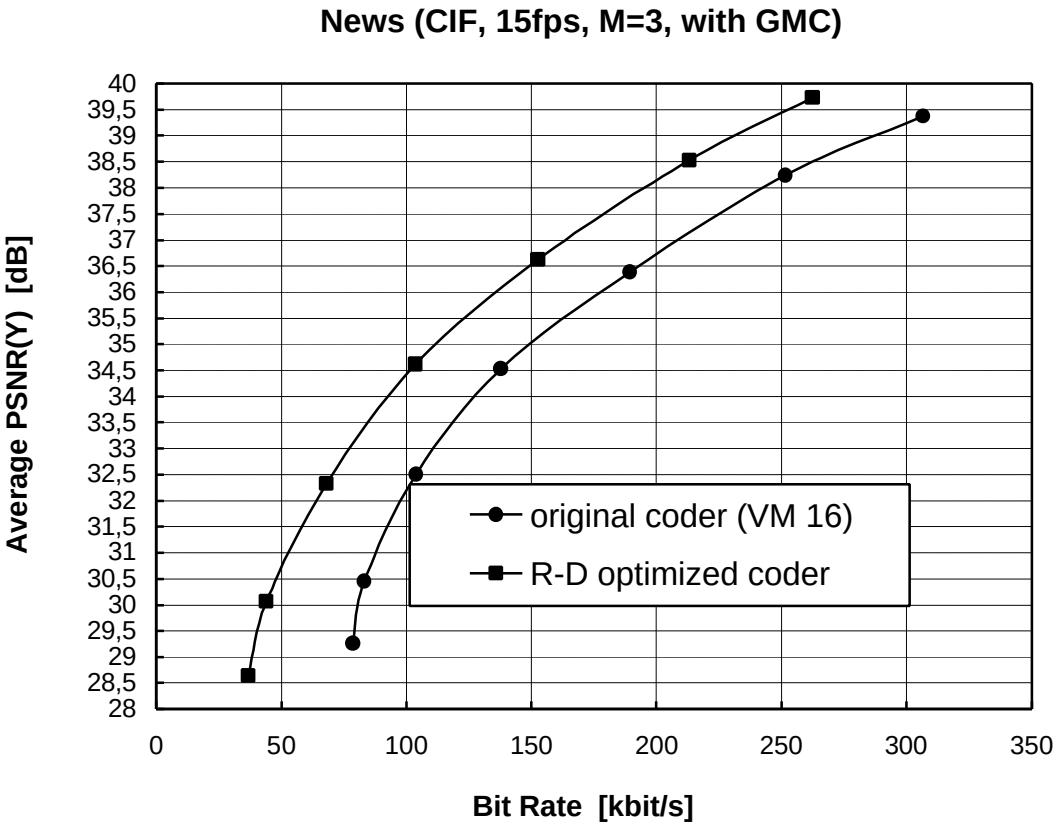



News (QCIF, 10fps, M=3, with GMC)
Average PSNR(Y) [dB]
Bit Rate [kbit/s]
original coder (VM 16)
R-D optimized coder
39
38,5
38
37,5
37
36,5
36
35,5
35
34,5
34
33,5
33
32,5
32
31,5
31
30,5
30
29,5
29
28,5
28
27,5
27
26,5
26
0
20
40
60
80
100


Tempete (QCIF, 10fps, M=3, with GMC)
Average PSNR(Y) [dB]
Bit Rate [kbit/s]
original coder (VM 16)
R-D optimized coder
36
35,5
35
34,5
34
33,5
33
32,5
32
31,5
31
30,5
30
29,5
29
28,5
28
27,5
27
26,5
26
25,5
25
24,5
24
0
50
100
150
200


Bus (CIF, 15fps, M=3, with GMC)
Average PSNR(Y) [dB]
37
36,5
36
35,5
35
34,5
34
33,5
33
32,5
32
31,5
31
30,5
30
29,5
29
28,5
28
27,5
27
26,5
26
25,5
25
0
250
500
750
1000
1250
1500
Bit-rate [kbit/s]
original coder (VM 16)
R-D optimized coder


Flowers & Garden (CIF, 15fps, M=3, with GMC)
Average PSNR(Y) [dB]
36
35,5
35
34,5
34
33,5
33
32,5
32
31,5
31
30,5
30
29,5
29
28,5
28
27,5
27
26,5
26
25,5
25
24,5
24
23,5
23
0
250
500
750
1000
1250
1500
Bit-rate [kbit/s]
original coder (VM 16)
R-D optimized coder


News (CIF, 15fps, M=3, with GMC)
Average PSNR(Y) [dB]
40
39,5
39
38,5
38
37,5
37
36,5
36
35,5
35
34,5
34
33,5
33
32,5
32
31,5
31
30,5
30
29,5
29
28,5
28
0
50
100
150
200
250
300
350
Bit Rate [kbit/s]
original coder (VM 16)
R-D optimized coder