CONSTANT QUALITY CONTROL BASED ON TEMPORAL DISTORTION BACKPROPAGATION IN HEVC

Médéric Blestel, Julien Le Tanou, Michael Ropert
Video Innovation Research, MediaKind

Overview

- Context and motivations:
  - Growing demand of the video traffic prompted more flexible platforms for delivery.
  - CBR is the most common rate control technique but suffers from several limitations:
    - Variable quality depending on the content complexity.
    - Over bandwidth consumption when delivering easy contents.
  - Quality-based Rate control algorithm should widely outperform CBR in terms of R-D performances, subjective experience and bandwidth savings.
- Contributions:
  - Inverting the paradigm:

From

$$\{QStep_k\}_{k=1}^{N_b} = ARGMIN(D_{Tot}), \quad subject\ to \sum_t \sum_{i_t} R_{i_t} = R_{Tot}$$

Into

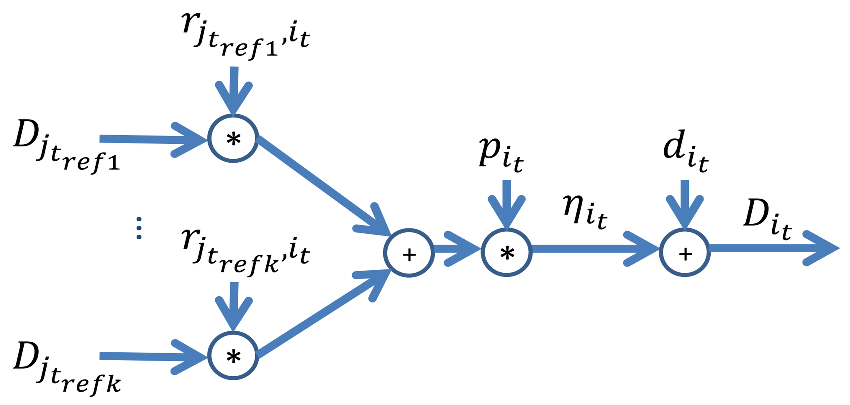
$$\{QStep_k\}_{k=1}^{N_b} = ARGMIN\left(\sum_t \sum_{i_t} R_{i_t}\right)$$

$$subject\ to \begin{cases} R_{Tot} < R_{Target} \\ \dfrac{\sum_t \sum_{i_t} D_{i_t}}{N \cdot T} = D_\mu \\ \forall i_t\ D_{i_t} < D_{Max} \end{cases}$$

  - Constant Quality Control (CQC) algorithm minimizes the bitrate under constraints:
    - A target video quality level.
    - A capped bitrate.
  - CQC reuses a temporal distortion propagation model to compute optimal local CUs quantizer.
- Outcomes:
  - -7.6% BD-BR PSNR improvement in average over state-of-the art algorithm in HEVC.
  - Meet the target level of quality with an average deviation of 6.7%.

Temporal Distortion Propagation Model

- Distortion propagation at CU level:



$$D_{i_t} = \eta_{i_t} + d_{i_t}$$

$$\eta_{i_t} = p_{i_t} \cdot \sum_{j_{t_{ref}} \in Ref(i_t)} r_{j_{t_{ref}},i_t} \cdot D_{j_{t_{ref}}}$$

  - i_t the CU index i in the frame numbered t
  - $Ref(i_t)$ the set of reference CUs used for motion compensation
  - p_{i_t} probability that the CU $_{i_t}$ is inter coded
  - r_{j,i_t} ratio of spatial overlap after motion compensation
  - η_{i_t} the projected distortion onto CU $_{i_t}$
  - d_{i_t} the intrinsic distortion of CU $_{i_t}$
  - $D_{i_t} = d_{i_t} + \eta_{i_t}$ the total CU distortion
- Generalizing propagation along a group of pictures (GOP) of length T:

$$D_{Tot} = \sum_{t=0}^{T-1}\left(\sum_{i_t} p_{i_t} \sum_{i_{t-1} \in Ref(i_t)} r_{i_{t-1},i_t}\left(p_{i_{t-1}} \sum_{i_{t-2} \in Ref(i_{t-1})} r_{i_{t-2},i_{t-1}} \dots \right.\right.$$
$$\left.\left. \dots\left(\dots p_{i_1} \sum_{i_0 \in Ref(i_1)} r_{i_0,i_1} d_{i_0} + d_{i_1}\right) + \cdots\right) + d_{i_t}\right)\Bigg)$$

Local Quantization Optimization Problem

- Find the set of local quantizers q_{i_t} of a GOP such:

$$\{q_{k_\tau}\}_{k_\tau \in Idx} \cup \{\lambda^*\} = ARGMIN\left(\underbrace{R_{Tot} + \lambda\left(D_{Tot} - \sum_t \sum_{i_t} D_{i_t}\right)}_{J_{Tot}}\right)$$

- Backward temporal distortion derivative:
  - Assuming that the intrinsic distortion (d_{k_τ}) related to a spatial position only depends on its local quantizers: $\dfrac{\partial D_{Tot}}{\partial q_{k_\tau}} = \dfrac{\partial d_{k_\tau}}{\partial q_{k_\tau}} U_{k_\tau}$
  - U_{k_τ} is obtained by the backpropagation of the U_{k_τ} values from the last non-reference frame of the GOP onto the first image, and defined by the recursion:

$$U_{k_{\tau-1}} = \sum_{i_\tau} p_{i_\tau}\, \rho_{i_{\tau-1},i_\tau}\, U_{i_\tau} + 1 \ and\ U_{n_{T-1}} = 1$$

$$\text{with } \rho_{j_{t-1},i_t} = \begin{cases} 0 & if \quad j_t \notin Ref(i_t) \\ r_{j_{t-1},i_t} & if \quad j_{t-1} \in Ref(i_t) \end{cases}$$

  - U_{k_τ} is an accumulation factor that is dependent on neither the distortion nor the rate
  - It is a weight related to how important is the current CU for coding the next frames in the GOP.
- After some mathematical developments, we obtain the optimal CU distortion D_{k_τ}:

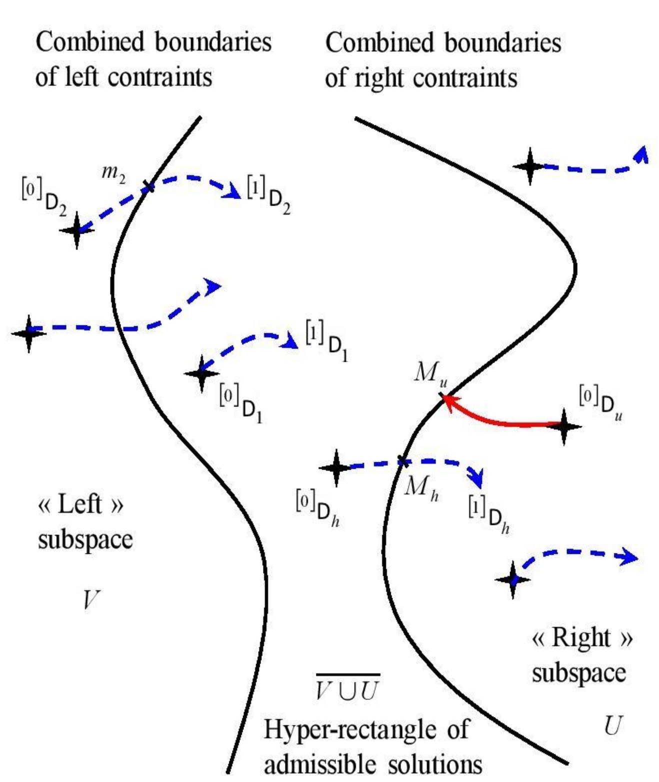
$$D_{k_\tau} = D_{Target} \cdot \frac{{U_{k_\tau}}^{-1}}{\sum_{t=0}^{T-1} \sum_{i_t} \cdot {U_{i_t}}^{-1}}$$

- Introducing a D to Q approximation based on a Laplacian distribution of residues $D_{k_\tau} = \dfrac{{\sigma^2}_{k_\tau} {Q^2}_{k_\tau}}{12{\sigma^2}_{k_\tau} + {Q^2}_{k_\tau}}$, optimal local quantizers are:

$$qp_{k_\tau} = 4 + 3\left(log_2(12) + log_2\left(\frac{{\sigma^2}_{k_\tau} D_{k_\tau}}{{\sigma^2}_{k_\tau} - D_{k_\tau}}\right)\right)$$

Optimization with inequality constraints

- When minimizing the problem with equality constraints, nothing guarantees that optimal quantizers do not infringe the CU maximal constraint and the maximal rate constraint.



- Step 1: Dealing with maximal CU constraint:
  - All violated distortion constraints are solved by packet thanks to the equality constraint.
  - Minimization is repeated until no constraints are violated.
- Step 2: Dealing with maximal rate constraint:
  - R_{Tot} is computed using the Shannon model:

$$R_{Tot} = -\frac{1}{2} . \sum_t \sum_{i_t} log_2\left(\frac{D_{i_t}}{c.{\sigma^2}_{i_t}}\right)$$

  - $If\ R_{Tot} > R_{Target}$, target distortion D_{Target} is rescaled.

$$D'_{Target} = D_{Target} . 2^{\frac{-2}{N.T}(R_{Tot} - R_v - R_{Target})}$$

Distortion Model Accuracy

- Using Machine learning techniques, target distortion is modelled and reached according to content characteristics:
  - From look-ahead estimations, each GOP target distortion is scaled.
- Evaluation process:
  - X265 open source software
  - Target GOP distortion is set on a PSNR base
  - For each target distortion, average MSE over the whole clip is measured
- Fairly efficient model:
  - Average deviation of 6,9%
  - Deviation increases with distortion
- Future works:
  - Refine R-D model to improve model accuracy
  - Take into account skipped-CU proportion

Test Sequences		Target Distortion (MSE)				
		10	20	30	40	50
Class B	Average	11,9	21,5	35,4	46,7	57,3
Class C	Average	9,0	18,4	28,5	38,9	53,8
Class D	Average	9,5	20,1	31,4	42,9	54,3
Class E	Average	12,7	21,9	28,9	47,0	62,2
All	Average	10,5	20,4	31,3	43,8	56,7
	Min	6,9	15,3	23,9	31,8	40,8
	Max	16,9	26,9	46,5	71,8	83,6
	Av. error (%)	5,3%	1,9%	4,3%	9,6%	13,5%

R-D Performances

- Experiments:
  - X265 open source software
  - Constant Qp evaluation
  - Followings JCT-VC Recommendations
  - YUV PSNR based results
- Results:
  - 7,6 % bandwidth savings
  - Up to 9,6% for low rates
- Conclusion:
  - Successful approach
  - Significant bandwidth savings particularly adapted to current OTT-TV purposes
- Future works:
  - Improve look-ahead estimations
  - Address psycho-visual based distortion

Test sequences		PSNR Based BD-BR	
		Full Rates (QP 22-42)	Low Rates (QP 27-42)
Class B	Average	-6,5%	-8,4%
	Best	-12,7%	-13,3%
	Worst	0,9%	-2,5%
Class C	Average	-9,3%	-12,6%
	Best	-19,5%	-21,6%
	Worst	1,3%	-0,1%
Class D	Average	-8%	-9%
	Best	-13,8%	-16,4%
	Worst	-4,9%	-6%
Class E	Average	-6,9%	-8%
	Best	-9,9%	-11,6%
	Worst	-1,9%	-2,2%
All	Average	-7,6%	-9,6%
	Best	-19,5%	-21,6%
	Worst	1,3%	-0,1%