

VVCにおける 4K8K 映像用 RD 最適化モデルの検討

A Study of RDO Model for 4K/8K Videos in VVC

近藤 雄一[†] 杉藤 泰子[†] 市ヶ谷 敦郎[†]
 Yuichi Kondo Yasuko Sugito Atsuro Ichigaya

1. はじめに

VVC (Versatile Video Coding)は、2020 年に国際標準化された最新の映像符号化方式であり、2018 年 12 月に開始した新 4K8K 衛星放送で採用されている HEVC (High Efficiency Video Coding)に比べて、符号化効率約 2 倍の改善を達成した符号化方式である^[1]。現在、総務省情報通信審議会では地上放送高度化に関する研究が進められており、衛星放送に比べて伝送帯域が狭い地上放送において高画質に映像を伝送するために VVC の利用が検討されている。

符号化制御技術の 1 つである RD (Rate-Distortion)最適化は、目標ビットレートにおいて歪みを最小化する最適な符号化パラメータを決定するための処理であり、符号化画質を左右する重要な技術である。しかし、VVC の参照ソフトウェア VTM (VVC Test Model)に用いられている RD 最適化手法は、比較的符号化難度の低い映像でしか評価されていない。放送される映像は符号化難度の低い映像ばかりではないため、放送を想定した条件下では既存の手法の性能が十分でない可能性がある。そこで、本稿では、より実用的な映像における符号化性能を検証するために符号化難度の高い映像を用いてレートと歪みの関係を分析するとともに、より適した RD 最適化モデルを新たに提案する。提案モデルを符号化制御に組み込むことで、演算量およびレート制御の精度を変えずに画質改善できることを示す。

2. VVC における RD 最適化

本章では、VTM における RD 最適化について説明する。

RD 最適化とは(1)式で表されるように、目標レート R_t において符号化歪み D を最小化する最適な符号化パラメータ (QP, 予測モード等)を選択することによって画質を最大化する処理である^{[2][3]}。

$$\min D \text{ s.t. } R \leq R_t \quad (1)$$

ラグランジュ未定乗数 λ を用いて(1)式は、 $\min(D + \lambda R)$ と書け、 $L = D + \lambda R$ をコスト関数と呼ぶ。このとき λ は、

$$\lambda = -\frac{\partial D}{\partial R} \quad (2)$$

によって求まる。 λ は映像によって異なる値となるが、符号化する映像の歪み D とレート R の関係 (RD 関係式) が分かっているならば(2)式から導出することができる。得られた λ を用いてコスト関数を最小化する符号化パラメータを選択することで最も効率的な符号化を実現できる。

VTM の RD 最適化では、レート R に基づいて λ を導出し、符号量を調整するための量子化パラメータ QP を決定する R - λ モデルが用いられている。当該モデルでは、RD 関係式として双曲線 (hyperbolic) モデルが仮定されており、歪み D を平均二乗誤差 (MSE) として RD 関係式は、

$$D = cR^{-k} \quad (c, k > 0) \quad (3)$$

[†] NHK 放送技術研究所 NHK Science & Technology Research Laboratories

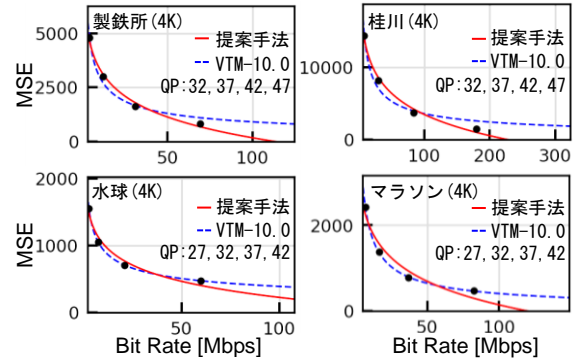


図 1 4K 映像を固定 QP で符号化したときの RD 曲線と表されている。したがって λ は(2)式より、

$$\lambda = -\frac{\partial D}{\partial R} = ckR^{-k-1} \quad (4)$$

である。ある i フレームの λ を λ_i とし λ_i で符号化したときのレート R_i と歪み D_i が分かると(3)、(4)式を解くことでフレームごとに異なるパラメータ c_i , k_i を求めることができる。次のフレームを符号化するには得られた c_i , k_i を $i+1$ フレーム目の c_{i+1} , k_{i+1} の予測値として用い、目標レート R_{i+1} を(4)式に代入することで λ_{i+1} を新たに導出し、コスト関数 L_{i+1} を最小化する符号化パラメータを決定する。ただし、QP ごとに L を最小化する符号化パラメータを決定すると計算量が膨大になるため、QP は(5)式に示す関係式を仮定して決定する^[4]。

$$QP = 4.2005 \times \ln \lambda + 13.7122 \quad (5)$$

ところで、双曲線モデルで示される(3)式が入力映像を実際に符号化した際の R と D の関係を適切に表していない場合、適切な λ を導出できないため、コスト関数を最小化する符号化パラメータが選択されても、符号化効率が最適化されない。そのためレートおよび画質が適切に制御されず、主観画質を低下させる可能性がある^[5]。つまり、RD 最適化では映像に適した RD 関係式の選択が重要である。

3. 4K8K 映像用 RD 最適化モデルの提案

本章では、まず 4K8K 映像の R と D の関係を分析し、符号化難度が高い場合に(3)式に示す双曲線モデルが適合しないことを示す。さらに、高符号化難度の映像に適した RD 関係式を新たに提案し、VTM に組み込んで効果を検証する。実験には、VTM-10.0 (2020 年 8 月発行)を使用した。

3.1 符号化難度の高い 4K8K 映像の RD 関係式

ITE 標準動画像^{[6][7]}の中で符号化難度の高い桂川・製鉄所・水球 (ゴール)・マラソン (スタート) をテスト映像 (表 1) として VTM Random Access で符号化したときの B ピクチャの実測した R-D 値に対する関数モデルとの相関の高さを調査した。テスト映像の符号化難度を VTM の性能評価に用いられている評価映像と比較したところ 2.4 倍～

21.7倍であった。符号化難度は B_{test}/B_{ctc} によって求めており、ここで、 B_{test} 、 B_{ctc} はそれぞれ、QP37のときの4Kテスト映像の符号量およびVTMの評価映像(Class A, 4種類)^[8]の平均の符号量である。値が大きいくほど符号化難度が高く、テスト映像の符号化難度が相対的に高いことがわかる。

R と D の関係は4K, 8Kそれぞれ900, 64フレームを用いて導出した。図1に示すように従来の双曲線モデルの場合、水球(難度2.4倍)・マラソン(難度5.1倍)は高精度に推定できたが、符号化難度が高い桂川(難度21.7倍)・製鉄所(難度8.0倍)では推定精度が低くなった。また表1に示すように、推定精度は解像度に依らず符号化難度が高い映像で低下する傾向があることが分かった。

そこで、符号化難度の高い桂川・製鉄所に着目して高精度に推定できるRD関係式を検討した結果、(6)式に示す対数モデルが適していることが分かった(図1, 表1)。

$$D = a \cdot \ln R + b \quad (a < 0, b > 0) \quad (6)$$

表1 符号化難度と近似曲線の決定係数

	4K			8K	
	難度	VTM	提案手法	VTM	提案手法
製鉄所	8.0	0.982	0.9923	0.9732	0.9978
桂川	21.7	0.9739	0.9921	0.9598	0.9991
水球	2.4	0.9998	0.9729	0.9992	0.9636
マラソン	5.1	0.9997	0.9442	0.9976	0.9608

3.2 対数モデルの実装

対数モデルを用いることによる画質改善効果を検証するためにVTMへの実装を行った。(2)式より λ は(7)式となる。

$$\lambda = -\frac{\partial D}{\partial R} = -\frac{a}{R} \quad (7)$$

ここでパラメータ a 、 b は、 λ_i 、 R_i 、 D_i を用いて(8)式で示される。ただし、対数モデルでは λ を求める変数は a だけであるため b の導出式を実装する必要がなく計算量が少ない。

$$a = \lambda_i R_i \quad \text{and} \quad b = D_i - \lambda_i R_i \ln R_i \quad (8)$$

1フレーム符号化する度に a を計算し、次のフレームの目標レート R_{i+1} を用いて(7)式から λ を求め、コスト関数 L を最小化する符号化パラメータを導出する。

符号化構造はLow Delay P^[9]とし、目標レートは現行の放送レートを考慮して10-80Mbpsの範囲から4点選択した。500フレーム符号化し、解析は符号化制御が安定する201フレーム目以降、300フレーム分を用いた。

3.3 符号化結果の評価

対数モデルによるRD最適化を用いたVTMで符号化した結果、双曲線モデルと同様に符号化制御が破綻することなく目標レートで符号化することができた(図2)。さらに、VTMに比べて同一画質における符号量(BD-Rate)を最大4.05%改善することができた(表2)。これは、適切な λ を用いてコスト計算が行われ、最適な符号化パラメータが選択されたことによる効果であると考えられる。

また、画質の時間変動が提案手法では小さくなることが確認できた。図3に桂川4Kの結果を示すが、その他の映像の結果も同様であった。ただし8フレーム毎に生じる大きな変動はシーケンス全体で画質を最大化するための符号化制御処理であり、予めVTMに実装されているものである。画質の時間変動が抑制されることで主観画質の改善が

予想され、実際に、専門家5名による簡易的な画質確認を行ったところ桂川4Kでは提案手法の方がVTMに比べて符号化劣化によるフリッカ状の劣化が抑制されており、その他の映像では視覚的な差がないことが確認できた。

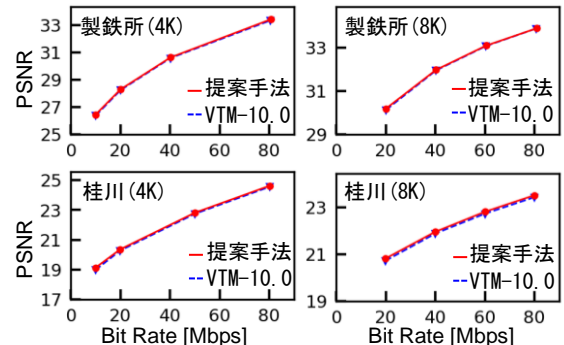


図2 Bit Rateと輝度信号のPSNRの関係

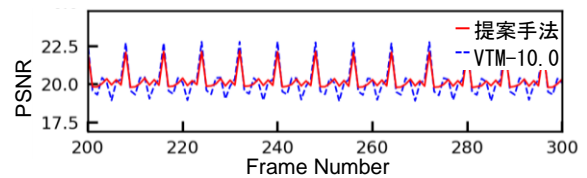


図3 桂川4Kのフレーム毎の輝度信号のPSNR

表2 輝度信号のBD-Rate

	4K	8K
製鉄所	-1.68%	-1.11%
桂川	-3.16%	-4.05%

4. おわりに

本稿では、4K8K高符号化難度の映像に着目してRD最適化モデルを修正することで、計算負荷を増やすことなく画質が改善することを示した。またフレーム間の画質の変動が減少し、主観画質改善の可能性も示唆している。

今後、双曲線モデルと対数モデルを切り替える指標を検討し、符号化難度に依存せず高効率で符号化できるようにRD最適化モデルの改善を進める予定である。

参考文献

- [1] "Report on VVC compression performance verification testing in the SDR UHD Random Access Category," JVET-T0097, Oct. 2020.
- [2] B. Li et al., "λ domain rate control algorithm for High Efficiency Video Coding," IEEE Trans. Image Process., vol. 23, no. 9, pp. 3841–3854, Sep. 2014.
- [3] B. Li et al., Adaptive Bit Allocation for R-Lambda Model Rate Control in HM document JCTVC-M0036, Apr. 2013.
- [4] B. Li et al., "QP Refinement According to Lagrange Multiplier for High Efficiency Video Coding," in Proc. Int. Symp. Circuits Syst. (ISCAS), Beijing, China, 2013, pp. 477–480.
- [5] Miaohui Wang et al., "Low-Delay Rate Control for Consistent Quality Using Distortion-Based Lagrange Multiplier," IEEE Trans. Image Process., vol. 25, no. 7, July 2016.
- [6] 映像情報メディア学会, 電波産業会, "超高精細・広色域標準動画像—Aシリーズ解説書" (2016)
- [7] 映像情報メディア学会, 電波産業会, "超高精細・広色域標準動画像—Bシリーズ解説書" (2017)
- [8] F. Bossen et al., "JVET VTM common test conditions and software reference configurations for SDR video," JVET-T2010, Oct. 2020.