

JVT符号化におけるレート制御に関する一検討

高木 幸一[†] 滝嶋 康弘[†] 中島 康之[†]

[†]株式会社KDDI 研究所 〒356-8502 埼玉県上福岡市大原 2-1-15

E-mail: [†]{ko-takagi, takisima, nakajima}@kddilabs.jp

あらまし 本稿では、JVT符号器におけるレート制御方式に関する検討を行ったので報告する。一般的に、動画符号化において特に固定ビットレートでの伝送を行うためにレート制御は必要不可欠な技術である。ところが、現在、JVTにおいてテストモデルに関する検討は始まったばかりであり、さらにこれまで符号化効率検証用に使われてきたテストモデルは固定量子化を基本としたもので、いわゆるレート制御に関する検討はまだほとんど行われていない。そこで、本稿では、JVT符号器における新たなレート制御方式を検討し、現在テストモデルで使用されているものとの比較を行う。

キーワード JVT符号化, テストモデル, レート制御

A Study of a Rate Control Scheme for a JVT coder

Koichi TAKAGI[†] Yasuhiro TAKISHIMA[†] and Yasuyuki NAKAJIMA[†]

[†]KDDI R&D Laboratories Inc. 2-1-15 Ohara, Kamifukuoka-shi, Saitama, 356-8502 Japan

E-mail: [†]{ko-takagi, takisima, nakajima}@kddilabs.jp

Abstract We report that a study of a rate control scheme for a JVT coder. Generally speaking, a rate control scheme is one of the most important techniques for a constant bit rate transmission. However, a rate control scheme for a JVT coder has hardly been studied. Therefore, we propose a new rate control scheme and compare it with conventional one.

Keyword JVT coder, Test model, Rate control

1. はじめに

ISO/IEC と ITU-T が共同で標準化を進めている JVT (MPEG-4 Part10 Advanced Video Coding(AVC)/H.264¹) は、符号化効率の観点で従来の方式より 50%符号量削減と言う、極めて高効率な符号化方式を目指しスタートした(詳細説明は[1][2]に譲る)が、近々の会合で設立されているアドホックグループの内容から、符号化の基本的な部分についてはほぼ仕様が凍結してきたと言える。

一方、それに伴い、検証用テストモデルに関する検討も始まりつつある。この中では、固定量子化パラメータに対し、すべてのモード、動きベクトル(MV)の中から最適なものを選択するための方式が定義されている。

ところが、エンコーダにおけるレート制御方式に関しては最近の標準化会合で提案が始められたところである。レート制御技術は固定ビットレート(CBR)での伝送を考えたときに極めて重要な技術であり、動画符号化におけるどの標準においても(非標準ではあるものの)検討が行われてきている。それゆえ、JVTにおいても同様にレート制御技術が必要である。

そこで本稿では、レート制御に関する検討を行うと共に、現在テストモデルに採用されているレート制御

¹ 本稿で対象とする符号化標準は MPEG-4 Part10 AVC, H.264, JVT(Joint Video Team)等いろいろな呼び方があり、これらは技術的に等価であるが、以下「JVT」と書くこととする。

方式との比較を行う。以下2節では、現在、JVT標準化各提案の検証用に使用されているテストモデルの概要を、3節ではそれを使ったレート制御手法を示す。一方、4節では、本稿で提案するレート制御手法を示し、5節で3、4節の方式を比較すべく、動画像を利用した計算機シミュレーションを行い、6節でこれらをまとめる。

2. テストモデルの概要

本節では、現在JVTで検討が進められている符号化テストモデル(非標準)について、動き推定とモード選択に絞って、その概要を示す。

本テストモデルには、比較的短時間で実行することを目的としたLow Complexity modeと、符号化効率を重視したHigh Complexity mode(Rate Distortion Optimization mode(以下RDOモードと書く))の2つのモードが存在する。

RDOモードにおける動き推定、およびモード選択は以下の手順で行われる。

1) 与えられた量子化パラメータ QP に対し、Lagrange乗数を以下のように与える。

$$\lambda_{MODE} = 0.85 \times 2^{QP/3}, \lambda_{MOTION} = \sqrt{\lambda_{MODE}}, \quad (1)$$

2) 複数の 4×4 イントラ予測 mode の中で cost 関数

$$J(s, c, IMODE | QP, \lambda_{MODE}) = SSD(s, c, IMODE | QP) + \lambda_{MODE} R(s, c, IMODE | QP)$$

を最小にするモードを選択する。ここで、 s は原画像、 c は予測画像、 $IMODE$ は 4×4 イントラ予測 mode の集合である。

3) 原画像 s と予測画像 c の差分に2次元Hadamard変換を施したものの絶対値和の1/2(これをSATDと呼ぶ)が最小になる 16×16 イントラ予測モードを選択する。

4) 8×8 画素サイズのブロックを複数ブロックに分割した各モードに対し、動き推定、参照フレームの推定を行い、cost関数 $SSD + \lambda R(MV, REF)$ の値を最小とするものをそれぞれ選択する。

5) $16 \times 16, 8 \times 16, 16 \times 8$ 画素サイズのMBを複数ブロックに分割した各モードに対し、動き推定、参照フレームの推定を行い、下記のcost関数値を最小とするものをそれぞれ選択する。

$$J(REF, \mathbf{m}(REF) | \lambda_{MOTION}) = SA(T)D(s, c(REF, \mathbf{m}(REF))) + \lambda_{MOTION} \cdot (R(\mathbf{m}(REF)) - \mathbf{p}(REF)) + R(REF)$$

ただし、 $\mathbf{m}(REF), \mathbf{p}(REF)$ はそれぞれ参照画像 REF に対する動ベクトルおよびその予測値、 $R(\cdot)$ は Universal

VLC(UVLC)を使って計算されたビット量である。

6) Bフレームに関しては予測方向も考慮に入れ、cost関数が最小になる予測方向を選択。

$$J(PDIR | \lambda_{MOTION}) = SA(T)D(s, c(PDIR, \mathbf{m}(PDIR))) + \lambda_{MOTION} \cdot (R(\mathbf{m}(PDIR)) - \mathbf{p}(PDIR)) + R(PDIR)$$

$PDIR$ は予測方向を示す。

6) 以上2~5の各モードで選ばれたそれぞれのパラメータに対し、

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

を最小にするモード $MODE$ を選択する。なお、P-frameではSKIPモード、B-frameではDIRECTモードも考慮に入れる²。

以上より、モードの選択が行われる。

3. テストモデルにおけるレート制御手法³

本節ではレート制御方式に関し、テストモデルで採用された方式についてその概要を述べる。

ところで、一般的にマクロブロック(MB)レベルでレート制御を行う場合、量子化パラメータ QP の値を変動させる。一方、前節で述べたとおり、最適なモードを選択するために QP を変動させないといけない。この両方の要因が存在することから QP をMBレベルで変動させることが極めて難しい。

そこで、JVTテストモデルでは、MPEG-2の検証モデルTM5[4]におけるレート制御方式を利用して、下記のような方法が採用されている。

1. ピクチャレベルで目標符号量の割当を行う(TM5 rate controlのstep1に相当)。

2. 各マクロブロックに対し、以下の処理を行う。

2-1. QP 値を直前に符号化されたMBの QP 値とし、前節の方式を利用して、最適なモード、 MV を求める。

2-2. アクティビティ値を考慮し、仮想バッファ占有量を計算し、そこから当該MBに使用する QP を計算する(TM5 rate controlのstep2に相当)。

2-3. 求められた QP を使って、前節の方式をもとに、最適な MV 、モードを選択する(ただし、求め

² これらのモードの意味は従来のMPEG標準で使われているものと等価と考えて良い。

³ 本手法は2002年7月JVT第4回Klagenfurt会合時で採用されたものである。2002年10月JVT第5回Geneve会合で新たなレート制御モデルが採用される可能性があるが、執筆時にそれを加えることは難しいため、その比較検証は今後の課題とする。

られた QP 値が 2-1 と同じであれば、この処理は行われない。

ただし、この方式で最適 MV モードが選択されるとは限らない。そこで、以下、2-2、2-3 を繰り返すことにより、最適パラメータの設定を行う必要があると考えられる。一方、[3]では、この繰り返しを一度行うだけで、レート歪み特性の観点から最適な点にほぼ収束することを確認している。そこで本稿でもこの繰り返しは一回のみとする。

4. 提案レート制御手法

前節の方式では、最適 MV 探索およびモード判定を各 MB で最悪 2 度ずつ行わないといけない(2-1, 2-3)。ところが、この処理は符号化における時間の大半を費やす場所であるため、可能な限り 1 度で済ませることが望ましい。そこで、本節では、前節の方式における処理 2-1 を削除し、仮想バッファ占有量のみによるレート制御を可能とする方式を提案する。なお、以下において、添え字 type は I,P,B 各ピクチャタイプを表す。

1. ピクチャレベルで目標符号量の割当を行う (TM5 rate control の step 1 に相当)。

まず、画面複雑度 X_{type} を定義し、符号化されたピクチャタイプにより更新する。更新式は $X_{type} = S_{type} QP_{type}$ とする。ここで、 S_{type} は当該ピクチャの符号量、 QP_{type} は実際に求められた量子化パラメータ QP の平均である。なお、 X_{type} の初期値はそれぞれ $X_i = 155 \times \text{bit_rate} / 115$ 、 $X_p = 15 \times \text{bit_rate} / 115$ 、 $X_b = 5 \times \text{bit_rate} / 115$ とする。これを用いて、目標符号量 T_{type} は TM5 と同様に以下の通り求める。

$$T_i = \max \left\{ \frac{R}{1 + \frac{N_p X_p}{K_p X_i} + \frac{N_b X_b}{K_b X_i}}, \frac{\text{bit_rate}}{8 \times \text{picture_rate}} \right\}$$

$$T_p = \max \left\{ \frac{R}{N_p + \frac{N_b K_p X_b}{K_b X_p}}, \frac{\text{bit_rate}}{8 \times \text{picture_rate}} \right\}$$

$$T_b = \max \left\{ \frac{R}{N_b + \frac{N_p K_b X_p}{K_p X_b}}, \frac{\text{bit_rate}}{8 \times \text{picture_rate}} \right\}$$

ここで、 K_{type} は各ピクチャタイプ間の複雑度の違いを表す比例成分、 N_{type} は当該 GOP で符号化されていないフレームの数、 R は当該 GOP において符号化され

ていないフレームに割り当て可能なビット量を示す。

なお GOP の最初のピクチャを符号化する前に、

$$R = \frac{\text{bit_rate}}{\text{pic_rate}} \times N + R_{\text{prev}}$$

とする。ただし、 N は GOP に含まれるフレームの数、 R_{prev} は直前の GOP における R とする。また、符号化後 $R = R - S_{type}$ とする。

2. 各 MB に対し、以下の処理を行う。

2-1. 仮想バッファ占有量を計算し、そこから当該 MB に使用する QP を計算する (TM5 rate control の step2 に相当)。MB m における仮想バッファ占有量 d_m を

$$d_m = d_0 + B_{m-1} - \frac{T_{type} \times (m-1)}{\text{MB_cnt}}$$

と置く。なお、 B_{m-1} は当該 $m-1$ 番目の MB までの消費符号量、 MB_cnt は全 MB 数である。

また、 d_0 は直前に符号化された同一ピクチャタイプを持つフレームの最後に求められた仮想バッファ占有量を使用する。なお、 d_0 の初期値はピクチャタイプごとに $d_{i0} = 30 \times r / 52$ 、 $d_{p0} = K_p \times d_{i0}$ 、 $d_{b0} = K_b \times d_{i0}$ 、また、 $r = 2 \times \text{bit_rate} / \text{pic_rate}$ とする。ここで 52 とは全 QP の取りうる値の総数を示している (TM5 および 3 節の手法では 31)。

以上をもとに、量子化パラメータ QP を以下の通り定める。

$$QP = \frac{52 \times d}{r}$$

ただし、 $0 \leq QP \leq 51$ とし、範囲外の場合は 0, 51 にそれぞれクリップする。

2-2. 求められた QP を使って、2 節の方式をもとに、最適 MV、モードを選択する。

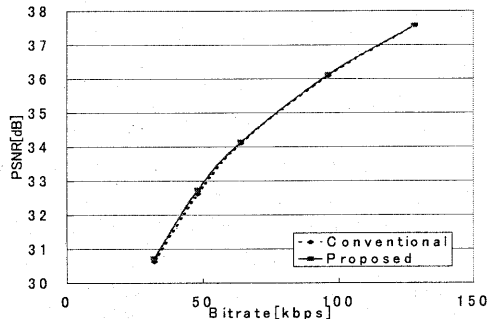
本方式では、3 節の方式と比較すると、2 節で述べられる最適 MV 探索、およびモード選択が一度しか行われず、その他の処理はほぼ等価であるため、確実に処理時間を短縮することが可能である。

5. 計算機シミュレーション

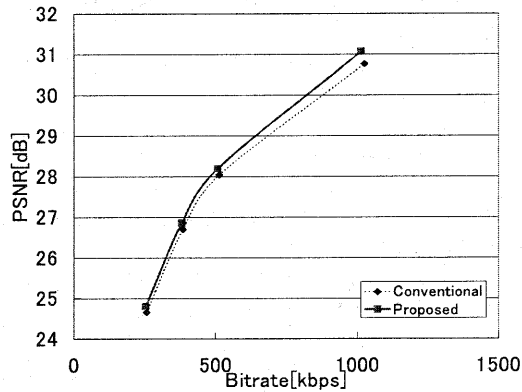
前節の提案手法の有効性を確認するため、テスト画像を使った計算機シミュレーション実験を行った。実験条件を表 1 に示す。

表 1. 実験条件

シーケンス	標準動画像 Foreman, Mobile&Calendar
エントロピー符号化	UVLC
動き探索範囲	32
MV精度	1/4画素単位
参照フレーム数	1
ピクチャ構造	フレーム構造のみ
RD optimization mode(2節参照)	On
GOP構造	GOP単位(N)=15frames, Pピクチャの間隔(M)は適宜



(a)Foreman(QCIF,10Hz,IPPP...,100frames,M=1)



(b)Mobile & Calendar(CIF,30Hz,IBBP..., 148frames,M=3)

図1. レート歪み特性

はじめに、両手法に対する、各ビットレートにおけるレート歪み特性の比較を図1に示す。破線が従来法、実線が提案法を示している。同図より、Bピクチャが存在している場合でも、そうでない場合でも、全ビットレートで提案手法の方が優れていることが確認できる。特にBピクチャが存在している場合、高レートでの向上が著しいことが確認できる。

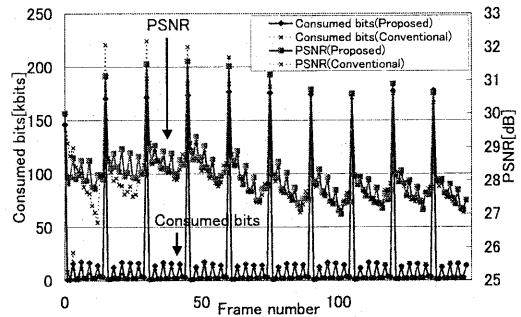


図2. 符号量とPSNRの推移
(Mobile & Calendar,512kbps)

次に、フレーム毎の符号量とPSNRの推移を図2に示す。実線が提案法、破線が従来法を示している。

同図より、特に従来手法は初期状態が不安定であるのに対し、提案手法は終始安定し続けていること、また、従来手法ではIピクチャで符号量を多く消費しており、同一GOPのPピクチャに悪影響を与えていることが伺える。

6. まとめ

本稿ではJVT符号化におけるレート制御に関し、従来法の問題点を指摘し、それを解決するための新たな方式を提案した。

動画像を使った計算機によるシミュレーションの結果、レート制御が終始安定して行えること、また、符号化時間の短縮のみならず、レート歪み特性の観点からも有効であることが確認できた。

今後の課題としては、あらゆる符号化モード(参照フレーム数、エントロピー符号化など)への適用可能性、および、極低ビットレートでの振舞い等に関する検証が挙げられる。

なお、本研究は、通信・放送機構より受託の「IPネットワーク上の音声・動画像配信(Webcasting)におけるQoS制御に関する研究開発」に基づき行われた。

文献

- [1] 鈴木輝彦, "MPEG-4 AVC|H.264の概要と標準化動向," 本研究会招待講演.
- [2] ISO/IEC JTC1/SC29/WG11 N4920, "Text of ISO/IEC 14496-10 FCD Advanced Video Coding" 2002.8.
- [3] Siwei Ma, et. al., "Rate control for JVT standard," JVT-D030, 2002.7.
- [4] ISO/IEC JTC1/SC29/WG11 N0400, "Test Model 5".1993.