

情報量予測型適応可変フレームレート制御方式
～超低ビットレート動画像符号化への応用を目的として～

宮地 悟史 松本 修一

miyaj@lab.kdd.co.jp

国際電信電話株式会社 研究所 画像通信グループ

〒 356 埼玉県上福岡市大原 2-1-15

動画像符号化, 特に 64kbps を下回る超低ビットレートにおいては, フレームレートの制御が, 画像品質や動きの滑らかさといったいわば動画像符号化の要となる性能を支配している. 本稿では, 符号化に先立って, 発生情報量をきめ細かく予測し, さらに動きベクトルの情報を用いることで, 画像品質を一定に保ちながら発生情報量を最小限に抑え, かつ視覚的に自然な動きとなるようなフレームレート制御方式について検討する. また本手法を, 背景分離型超低ビットレート動画像符号化方式に応用し, 量子化器の最適化と合わせて, その有効性について述べる.

A Method of Entropy Predictive Frame Rate Control

— for Very Low Bit-rate Video Coding —

Satoshi MIYAJI

Shuichi MATSUMOTO

KDD R&D Laboratories

2-1-15 Ohara Kamifukuoka-shi Saitama 356, Japan

The control of frame rate contributes effectively overall codec performance represented by the picture quality and motion smoothness in very low bit-rate video particularly less than 64kbit/s. This paper describes the method of frame rate control using precise entropy estimation and motion vector information to keep the picture quality high and get subjectively natural motion. And the result of applying the suggested method to VLBR codec with background separation and advanced quantization control is shown.

1 はじめに

近年、超低ビットレート動画符号化の研究が活発に行われており、これらは H.263 に代表される従来からの波形符号化技術に基づくものと、画像の内容について記述する知的符号化とに大別される。いずれの方式に関しても、64kbps を下回るような超低ビットレート領域においては、フレームレートの制御が、画像品質や動きの滑らかさといったいわば動画総合品質の要となる部分の良し悪しを決定する鍵を握っている。

本稿では、前者に属する超低ビットレート波形符号化方式に應用することを目的として、符号化に先だって、予め発生する情報量をきめ細かく予測し、さらに動きベクトルの大きさを考慮して、画像品質と動きの滑らかさとの関係を最適なポイントで動作させるフレームレート制御方式について述べる。

本方式では、H.263TMN5[2] で用いられている符号化後にフレームレートが決定される制御方式とは異なり、画像の情報量を考慮してフレームレートが決定されるため、与えられたフレームスキップ内で、画像品質を一定に保ちながら最大限の量子化が行える。また動きベクトルの大きさを考慮しているため、動きの大きな画像に対してもフレームスキップによる動きの不自然さを最小限に抑えることができる。

また、本フレームレート制御方式を筆者らが過去に報告した背景分離型超低ビットレート動画符号化方式[3]に應用し、発生情報量予測の際に得られる、マクロブロック毎の画像性質を十分に利用した量子化器最適化を合わせて行い、従来のバッファ制御情報が支配的であった量子化制御との相違について述べる。

2 適応フレームレート制御

2.1 H.263TMN5 での制御方式

H.263TMN5 でのフレームレート制御方式[2]を図1に示す。まず前フレームでの情報発生量を基にターゲットビット数を定め、量子化制御を行い、符号化する。その結果、バッファメモリの蓄積量から、ターゲットビット数と実際に発生したビット数との偏差を吸収するようにフレームレートの制御を行っている。

この方式では、例えば、画像の持つ情報量が前フレームに引続き多い場合、まず量子化ステップの増大となって現れ、画像の品質が低下する。

```

buff_content = buff_content + Bi;
while (buff_content > 3  $\frac{R}{FR}$ ) {
    buff_content = buff_content -  $\frac{R}{FR}$ ;
    frame_incr++;
}
R      : ビットレート
FR     : フレームレート
Bi   : 現フレームで発生したビット数
    
```

図 1: H.263TMN5 でのフレームレート制御

また、シーンチェンジが起これ、その発生情報量が前フレームと比較して大幅に増大した場合、前フレームを基にしたターゲットビット数による量子化制御では、実際の発生情報量との偏差が大きくなり、それがフレームレートにより吸収されるため、フレームスキップが大きくなり、その結果、シーンチェンジの度にフレームがスキップするといった現象となる。このように、フレームレートを符号化後に決定する場合、スキップ数の決定は量子化後の副産物的要素に過ぎず、画像の持つ情報量や動き量を反映することができない。

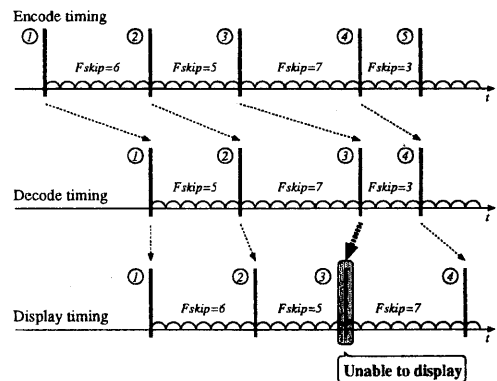


図 2: フレーム間隔の不一致

さらに、この方式では、図2に示すように、受信側で、エンコード時のフレーム間隔を維持しようとした場合に、図2中の No.3 フレームのように、表示すべきタイミング時に画像データが到達していないといったことが起これ得る。これに対処するためには、画像データが到達した時点で表示する、あるいは、あらか

じめ最大遅延量を規定し、No.1 フレームの表示を遅らせることが必要である。前者の場合は、フレームの間隔が送信側と受信側とで異なる点で好ましくない。また、後者については、H.263TMN5 でのフレームレート制御では、符号化後にフレームレートを決定する為、フレーム遅延の最大量が規定できない。

以下に、これら従来方式の問題点を解決するため、予め発生情報量や動き量を予測し、符号化前に最適なフレームレートを決定する手法について述べる。

2.2 情報量予測によるフレームレート制御

ここでは、フレームの持つ情報量を利用して、符号化前に次フレームまでのスキップ数を決定するフレームレート制御方式について述べる。本方式のブロック図を図3に示す。

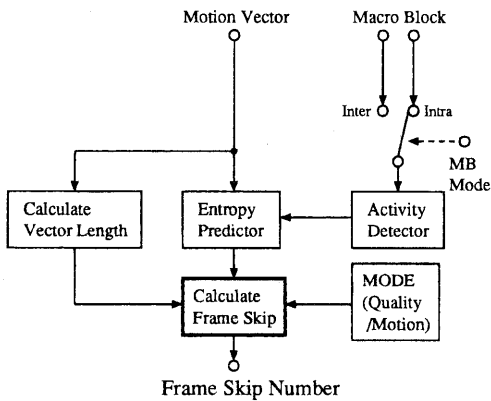


図 3: 情報量予測によるフレームレート制御

本制御方式では、符号化前の前処理部において、以下のパラメータによりフレームスキップ数を決定する。

- 予測発生情報量（ビット数）
- 符号化モード（画質優先/動き優先）
- 動ベクトルのスカラ長
- バッファメモリ占有状態

これにより、後述する画像の持つ情報量や動き量を十分に利用することができる。また、前もってフレームレートを決定するので、最大遅延量を規定ことができ、送信側と受信側とでのフレーム間隔の不一致を無くすることができる。

2.2.1 発生情報量の推定

画像の発生情報量は、アクティビティと相関があるため、その値から符号化前に発生情報量を推定することができる。本方式では、マクロブロック毎に次式で示される予測式を用いて発生情報量の推定を行う。

$$B_{mb} = fp(act) + B_{ex} \quad (1)$$

関数 fp はマクロブロックのアクティビティ act を引数とする量子化により発生する情報量の予測関数である。また B_{ex} は、動ベクトル及びその他の付加ビットの合計ビット数である。

上式より算出されたマクロブロックの予測発生情報量 B_{mb} は、次節で述べるフレームスキップ数の決定に用いられる。

2.2.2 フレームスキップ数の決定

まず、マクロブロック毎に、式(1)より予測発生情報量 B_{mb} を算出し、式(2)によりフレーム全体の総発生量を算出する。これがフレームスキップで過不足無く吸収できるように、基本スキップ数を決定する(式(3))。情報発生予測関数は、視覚的画質劣化が最小限となる中で、最大限のステップで量子化した場合の発生情報量を推定する関数とする。

式(3)において、画質/動き優先の各モードを考慮するため、画質優先モードについては画質/動きパラメータ $k_m = 1$ とし、基本スキップ数をそのまま用い、動き優先モードについては、 $0 < k_m < 1$ とし、動きが滑らかになるようにスキップ数を減少させる。

また、動きベクトルのスカラ長を考慮し、大きな動きが存在する場合には、フレームスキップ数を減少させる。これにより大きな動きを持つ画像に対して、視覚的動きの不自然さを解消できるだけでなく、フレーム間動き予測の効率化を図る。

さらにバッファメモリの占有状態により、何らかの原因で前フレームの量子化が効率良く行われなかった場合でも、フレームスキップ数に反映させ、現フレームの符号化制御に与える影響を最小限にする。

画面全体での発生情報量予測値を B 、マクロブロック毎の発生情報量予測値を B_{mb} 、画質/動き優先モード係数を $k_m (0 < k_m \leq 1)$ 、動きベクトルのスカラ長を $L_{mv} (0 \leq L_{mv} \leq 16\sqrt{2})$ 、バッファメモリ占有割合を $B_{MO} (0 \leq B_{MO} \leq 1)$ 、 α 、 β を定数として、基本ス

キップ数 F_S , および最終的なフレームスキップ数 F_{skip} を次式で与える.

$$B = \sum_{mb=1}^{MB} B_{mb} \quad (2)$$

$$F_S = \left(\frac{B}{\frac{BR}{30}} - 1 \right) k_m \quad (3)$$

$$F_{skip} = \left(F_S - \frac{L_{mv}}{\alpha} \right) \left(1 + \frac{B_{MO} - 0.5}{\beta} \right) \quad (4)$$

3 背景分離型超低ビットレートコーデックへの応用

過去に文献 [3] において, 背景分離型の超低ビットレートコーデックについて報告した. 今回, 前節で述べたフレームレート制御方式を新たに超低ビットレートコーデックへ応用し, さらに背景分離部, 量子化制御及びバッファ制御方式の再検討を行ったので報告する.

3.1 コーデックの全体構成

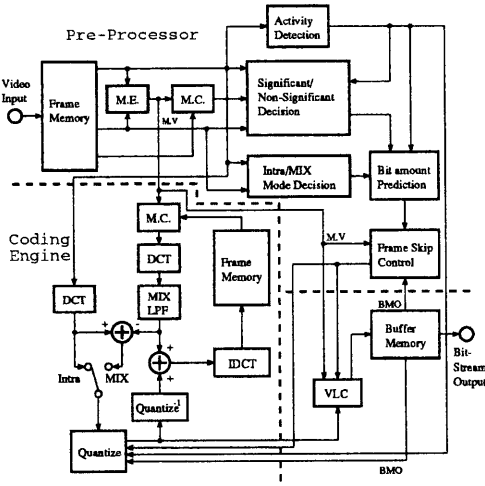


図 4: システムの全体構成

背景分離型可変フレームレート超低ビットレートコーデックのシステム全体構成を図 4 に示す. 本システムは, 前処理部, 符号化部, 及び伝送部の各部から構成される.

前処理部では, 動き検出背景分離 (有意/非有意判定) [3], 画像の絵柄変化度 (アクティビティ) 検出, 符号

化モード判定, 及び前述した発生情報量予測型フレームスキップ数決定の各処理を行う.

符号化部では, 前処理部で得られた情報を基に, ミックスモード符号化 [3] ならびに視覚特性を反映させた適応量子化制御を特徴とした動き補償フレーム間 DCT 符号化を行う. 本システムでは前処理において効果的な発生情報量予測を行っているため, 後述する適応量子化制御と合わせて, 発生するブロック歪みは最小限のものとなっている.

伝送部は, 可変長符号化器とバッファメモリから構成され, 画像データをビットストリームとして出力する. 本システムは可変フレームレートを採用しており, これを吸収するため, バッファメモリは最大遅延量に相当するビット数を蓄積できる容量を有するものとする. また, これにより前述した通りフレームスキップ数の送受信側での不一致を防ぐことができる.

以下に各部における主要機能について述べる.

3.2 動き検出背景分離

不必要な情報発生を避け, また視覚的に高品質な画像を得ることを目的とした動き検出背景分離を行う.

まず, 符号化雑音のない原画像同士で動き検出を行い, マクロブロック毎の動き予測誤差を求める. この時, 動きベクトルをその大きさが徐々に大きくなるように発生させ, 動き予測誤差の絶対値和 D が, 閾値 T_H を下回った場合, そのブロックを非有意ブロック (背景) と判定する. 非有意ブロックに関しては, 動きベクトルのみを送り, 一切の量子化を行わない. また, 非有意ブロックの動きベクトルが $(0, 0)$ の場合を静止ブロックとして取り扱い, 静止を示す ID を除いて一切の符号化を行わない.

以上のマクロブロック分類手順を図 5 に示す. また, 有意, 非有意, 静止の分類を表 1 に示す.

表 1: マクロブロックの分類

分類	ID	判定条件	伝送情報
静止	0	$D < T_H, MV=0$	なし
非有意	10	$D < T_H$	MV
有意	11	$D \geq T_H$	MC-DCT

閾値 T_H に関して, ブロックアクティビティ及び動き

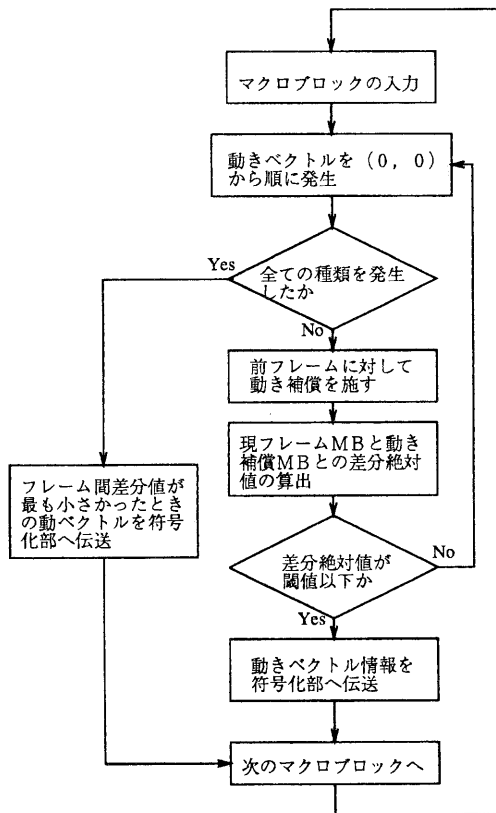


図 5: マクロブロック分類手順

ベクトルのスカラ長を考慮することとする。前者については、雑音のマスク効果を利用し、アクティビティの低い平坦なブロックは、量子化を行わないことの影響を受けやすいので、アクティビティの高い複雑なブロックに比べて閾値 T_H の値を小さくする。従って閾値 T_H は、アクティビティに関して単調増加の傾向とする。

また、後者については、動ベクトルを可変長符号化する場合、ベクトル長の小さいものは符号長も短くなるように割り当てられているため、動き予測誤差が同程度であれば、動ベクトル長ができるだけ小さくなるようにする必要がある。このため、 T_H は、動ベクトル長に関しては、単調減少となるようにする。

3.3 適応量子化制御

H.263TMN5[2]での量子化制御は、前フレームでの符号化結果から目標ビット数及び量子化ステップサイズを定め、それを用いて現フレームを実際に符号化し、発生した情報量との偏差から、量子化ステップを修正するものである。従って、視覚特性が全く考慮されておらず、また前フレームでの情報が基になっているため、動きの激しい画像や、シーンチェンジなどの場合に、適切な量子化制御が困難である。また、マクロブロック毎に量子化ステップが大幅に変化することがあり、この場合ブロック歪みが顕著に生じる。

これに対し、本コーデックでは以下の7パラメータによる適応量子化制御を行う。

- 係数位置 (空間周波数視覚感度) m, n
- 輝度 or 色差 YC
- 符号化モード mode
- 出力バッファ占有量 B_{MO}
- ブロックアクティビティ σ^2
- 予想フレーム発生情報量 B
- フレームスキップ数 F_{skip}

量子化ステップサイズ QP は次式で与えられる。

$$QP = H(m, n, YC, mode) \times f(Q_B(\sigma^2), F_{skip}, B, B_{MO}) \quad (5)$$

関数 H は、YC, mode をパラメータとする空間周波数視覚感度係数である。また Q_B は以下で述べる目標量子化ステップであり、 f は目標ステップ補正関数である。

目標量子化ステップサイズの決定は次のようにして行われる。画像の局所的変化度 (ブロックアクティビティ) により4種類 (Flat, Gradient, Edge, Detail) に分類する。各アクティビティの領域において、主観評価により劣化が許容できる最大ステップサイズを決定する。すなわち、これを用いて画面を一様に量子化した場合、画質劣化は最小限であり、かつ画面全体での各アクティビティ間の品質の連続性も保たれるように決められる。

本コーデックでは、フレームスキップ数決定のため、目標量子化ステップによる現フレームの発生情報量の

予測を行っている。従って、次の2つについて、目標量子化ステップを変更する必要がある。

- 予想発生量により決定されたフレームスキップ数が最大規定遅延量を越える、または予想発生量が1フレーム分の情報量を下回る場合
- 予想発生量と実際の発生量との食い違い

前者の場合、補正関数 f にパラメータ B, F_{skip} を与え、フレーム全体における目標量子化ステップを補正する。また、後者の場合には、マクロブロック毎にバッファ占有量 B_{MO} により、ステップサイズの更新を行う。

これに関して、本コーデックの場合には、情報発生量の予測を現フレームに対して直接行うことを特徴としている。従って、大きな動きや、シーンチェンジ等に対しても予想発生量が実際と大きく異なることは少なく、マクロブロック毎の量子化ステップ更新はほとんど行われず、このためブロック歪みは最小限のものとなっている。

3.4 バッファ制御

本コーデックでは、符号化の要となるフレームレート制御や量子化制御を、現フレームの予想発生情報量を基本として行っているため、予測した発生量と実際の発生量との食い違いをバッファメモリ占有量 (B_{MO}) から把握し、制御に反映させる。

B_{MO} は、全バッファ容量に対する現在バッファ内に蓄積されている量であるが、可変フレームレートの場合、フレーム間スキップ数が1枚の場合と10枚の場合とでは、次フレーム符号化までの時間間隔が異なるため、例えば、実際のバッファメモリ内の情報蓄積量が同じ場合でも、スキップ数1枚の場合に比べて10枚の場合の方がバッファメモリに余裕があることになる。

可変フレームレートに適応したバッファメモリを、図6(a)の2バッファ構成で考える。第一バッファは、フレームレートに応じて、可変ビットレートで出力されるものとし、これによりフレームレートによらず、バッファのスケールは一定に保たれる。また、第二バッファは、可変ビットレートで入力されたものを、一定レートで出力するためのものである。

従って、第一バッファの占有量を B_{MO} として用いる。

図6において、現フレームの処理を開始してからの時刻を t として、 $b(t)$ を図6(a) 第一バッファの蓄積

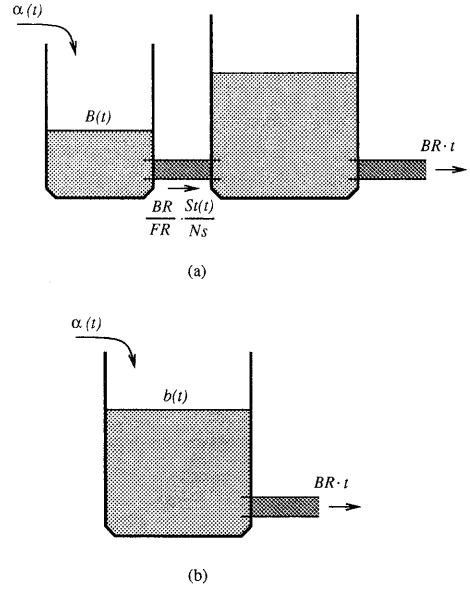


図6: 出力バッファ概念図

量, b_{max} を第一バッファの最大容量, $B(t)$ を図6(b) バッファの蓄積量, $b_s(t)$ を時刻 $T=0$ から $T=t$ までに発生した情報量, BR をビットレート, FR をフレームレート, N_s をフレーム中の全ストライプ数, $S_t(t)$ を現在処理中のストライプ番号とする。

図6(a)の第一バッファの占有量 $B_{MO}(t)$ を次式で定義する。

$$B_{MO}(t) = \frac{b(0) + b_s(t) - \frac{BR}{FR} \cdot \frac{S_t(t)}{N_s}}{b_{max}} \quad (6)$$

また、

$$B_{MO}(t) = \frac{b(t)}{b_{max}} \quad (7)$$

であるので、

$$B_{MO}(t) = \frac{b_s(t) - \frac{BR}{FR} \cdot \frac{S_t(t)}{N_s}}{b_{max}} + B_{MO}(0) \quad (8)$$

また、図6(b)のバッファ蓄積量 $B(t)$ は、コーデックの実際のバッファ蓄積量であり、

$$B(t) = B(0) + b_s(t) - BR \cdot t \quad (9)$$

である。式 (8) および式 (9) より、情報発生量 $b_s(t)$ を消去して次式を得る。

$$B_{MO}(t) = \frac{B(t) - B(0) + BR \cdot t - \frac{BR}{FR} \cdot \frac{S_i(t)}{N_s}}{bmax} + B_{MO}(0) \quad (10)$$

ただし、フレームレート FR とフレームスキップ数 F_{skip} との関係は、

$$FR = \frac{30}{1 + F_{skip}} \quad (11)$$

となる。よって、式 (10) より、実際の 1 バッファ構成のバッファ蓄積量から、2 バッファ構成を想定した時の第一バッファの B_{MO} を算出し、これをフレームレート制御、適応量子化制御に利用する。

4 シミュレーション実験

情報量予測型適応可変フレームレートを用いた超低ビットレートコーデックの、計算機シミュレーションにより H.263TMN5[2] との比較を行った。

4.1 実験条件

入力信号、前処理部、符号化部、伝送部各々に関する実験の条件を表 2~4 に示す。

表 2: 入力信号

画像形式	Y, C_r, C_b
画像サイズ (Y)	176 画素 × 144 ライン
(C)	88 画素 × 72 ライン
フレームレート	30 枚/秒
フレーム数	300 フレーム (10 秒)

4.2 実験結果

各フレームにおける PSNR を図 7 に示す。横軸が処理フレーム番号、縦軸が PSNR である。図中のプロットは、符号化処理の行われたフレームに対して行われており、各プロット間隔はフレームスキップ数を表している。

画像 RACE は、シーケンス前半は静止している背景の中を数台の自動車を通り抜けて行くシーンで、シー

表 3: 前処理, 符号化

フレーム間引き数	$0 \leq F_{skip} \leq 9$
有意 MB 判定閾値	$1.0 \leq T_H \leq 5.0$
動き検出	±16 画素, 1/2 画素精度
精細度分類	Flat, Gradient, Edge, Detail
符号化モード判定	Intra, MIX (C 信号は Inter)
量子化制御	7 パラメータ適応量子化
符号化レート	$BR=50.9\text{ kbit/sec}$

表 4: 伝送部

仮想バッファ容量	$bmax=32768\text{ bits}$
実バッファ容量	32768bits
エラー訂正	RS(63,59)
冗長度	9.253%
伝送レート	64kbit/sec

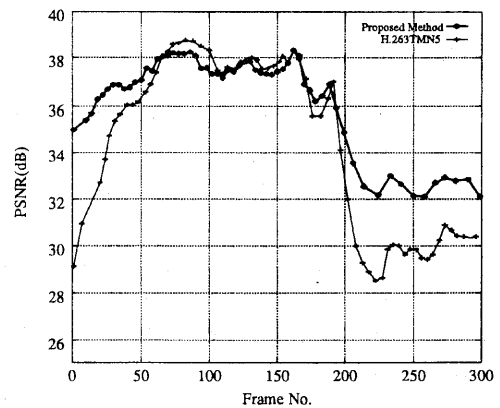


図 7: 各フレームにおける PSNR (RACE)

ケンス後半(約 190 フレーム以降)で、一台の車を追いかけられるようにカメラがパンニング動作をし、各フレーム毎に画像全体が大きく変化するものである。

図 7 によると、まずシーケンス先頭から 70 フレーム前後まで、提案方式の PSNR 値が高い。本方式は 1 枚目の Intra Picture に比較的多くのビット数を割り当てている。この為、1 枚目から 2 枚目にかけてのフレームスキップは、H.263TMN5 に比べて大きくなっているものの、フレームの符号化が高品質に行われているので、背景分離が有効に作用し、かつ、有意部分の動き補償が効率良く行われているため、フレームレートはどちらの方式もほぼ同様でありながら、PSNR の値が大幅に向上している。

また、注目すべきは、シーケンス後半 190 フレーム以降のカメラが車を追いかける情報発生が多いシーンである。H.263TMN5 では量子化ステップの増大によるブロック歪みが生じ、また視覚特性が考慮されていない為、道路面等比較的フラットな領域における品質劣化が目立つ傾向にある。一方、提案方式では、情報量予測によるフレームスキップ数制御により、品質の劣化を最小限にし、またこれによるフレームスキップ数の増大も視覚的に問題になる程ではない。さらに発生情報量を予測したフレーム全体での量子化ステップ制御により、個々のマクロブロック毎の量子化ステップ更新がほとんど行われていない為、視覚特性を十分に考慮した目標量子化ステップが生かされ、ブロック歪みが極めて小さくなっている。

5 まとめ

符号化前にフレームで発生する情報量を予め予測し、適応的にフレームレートを制御する方式について検討し、これを背景分離型超低ビットレートコーデックに応用した。

従来方式では、前フレームでの発生情報量を基に量子化制御を行い、その結果からフレームスキップ数を決定していた。これに対し提案手法では、発生情報量予測に基づき符号化前にフレームレートを決定するため、フレームの持つ特徴を最大限に利用することができる。その結果、画像品質を限りなく一定に保ち、かつ動き情報も考慮して、視覚的に動きの滑らかな画像が得られた。

超低ビットレートコーデックへの応用に関して、フ

レームの特徴を利用した効果的なフレームレート制御により、量子化制御でのブロック毎のステップサイズ更新が最小限となり、さらに視覚特性を十分考慮した 7 パラメータによる量子化制御によりブロック歪みが大幅に低減した。

今後は、視覚特性のさらなる改善や注視点情報の利用について検討を行う予定である。

謝辞

日頃御指導頂く国際電信電話(株)研究所の村上所長に感謝いたします。

参考文献

- [1] ITU-T Study Group 15 : “DRAFT RECOMMENDATION H.263; Video coding for narrow telecommunication channels at < 64kbit/s”, March 1995
- [2] ITU-T Study Group 15 : “VIDEO CODEC TEST MODEL, TMN5”, January 1995
- [3] 宮地, 松本 : “背景分離型適応可変フレームレート動画像符号化方式”, 情処研報, 96-AVM-12, Vol.96, No.17, pp.13-20 (1996)