

超低ビットレート動画像符号化方式
～空間的，時間的，統計的情報量制御による高品質符号化～

宮地 悟史 松本 修一

miyaji@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

64kbps 以下をターゲットとした超低ビットレート動画像符号化においては，中～高ビットレートで一般的に用いられているバッファ状態の帰還による量子化制御では，過去の発生情報量に頼らざるを得ない，画像サイズが小さいためマクロブロック単位の帰還制御のみではブロック歪みが目立つ，等といった点で最適な制御とはいえない．今回筆者らは画像の空間的情報量制御，時間的情報量制御，統計的情報量制御の3制御を組み合わせた新たな画像符号化方式を提案し，各々の制御の詳細，及び相互の制御関係について述べる．また，本方式により実際の超低ビットレートコーデックシステムを構築し，その符号化性能について検証する．

A Method of Very Low Bit-rate Video Coding
— with Spatial, Temporal and Entropy Coding Control —

Satoshi MIYAJI Shuichi MATSUMOTO

KDD R&D Laboratories
2-1-15 Ohara Kamifukuoka-shi Saitama 356, Japan

The quantization control by the status of buffer memory, which is conventionally used for the high bit-rate video coding, is not the best way for the very low bit-rate video less than 64kbit/s in respect of depending on the number of bits previously spent, and the large macroblock size (16x16) for QCIF. This paper proposes the new video coding method by spatial, temporal and entropy control. The details of each control and the relationship among these controls is described. The very low bit-rate video coding scheme based on the proposed method and the performance of this system is also shown.

1 はじめに

ISDN回線、PHSの急速な普及に伴い、32k、64kbpsといった超低ビットレートでの移動体画像通信、テレビ会議、監視カメラ映像伝送等への応用を目的とした高品質画像符号化が注目されている。この分野においては、ITU-T勧告H.263が画像品質の点で比較的高い評価を得ている。

しかし、H.263では、符号化後に生じる情報発生に基づく制御のため、前もって画像の性質を把握することができず、バッファ蓄積量の帰還が支配的となる量子化制御に依らざるを得ない。このため、符号化が易しい領域での情報の隙間を、符号化の難しい領域が有効に使うことが出来ず、効果的な符号化制御が行えない。また、従来中～高レートで一般的に用いられてきたこの制御方式は、比較的小さな画像を取り扱う超低ビットレート符号化では、画面における1つのマクロブロックが大きなものとなるため、きめ細かい制御が行えず、ブロック歪み等の原因となる。

さらに、リアルタイム双方向通信を意識したときの遅延制御が考慮されておらず、可変フレームレートで伝送する際、送信側と受信側でのフレーム間隔が一致しないといった現象が生じる。

これに対し、筆者らは、空間的情報量制御、時間的情報量制御、統計的情報量制御の3制御を組み合わせた新たな動画像符号化方式を提案する。空間的情報量制御として動き補償背景分離を、時間的情報量制御としてフレームレート制御を、統計的情報量制御として適応量子化制御をそれぞれ用いる。本稿では、各制御の詳細について述べ、これら3制御から成る互いの制御関係及び動作ポイントの最適化について検討する。

また、実際のハードウェアでの実現を考慮したシステム構成を示し、パラメータ最適化について述べ、従来方式(H.263TMN5)との性能比較を行う。

2 システムの制御

2.1 従来の制御系

図1に、従来の動画像符号化方式として、ITU-T H.263TMN5の制御系を示す。本図は、量子化に帰還される制御情報に着目して描かれた概念図であり、動き検出、モード判定部等は省略してある。

従来の符号化方式では、動き予測、モード判定がなされたマクロブロックが、符号化部に入力され、量子化が行われる。量子化により発生した情報は、バッファメモリに蓄積され、過去の発生量や理想発生量との比較を行い、量子化ステップサイズ変更へと帰還される。1フレームの符号化が終了した後、さらにバッファ残量と理想値との比較を行い、その偏差をフレームレ

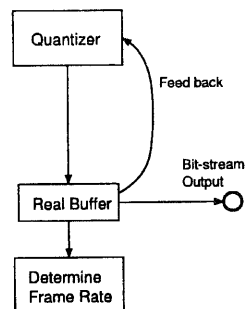


図1: 従来の制御系 (H.263TMN5)

ト変更にて吸収している [2]。

2.2 従来方式の問題点

従来方式では、画像の性質を前もって把握することができない為、次の様な問題を生じる。

- 情報発生の大いシーンにおいて
 - 量子化ステップの増大が起こり、画像品質の低下を招く
 - 理想発生量や前フレームの発生量に基づく制御のため、特にシーンチェンジ等で画像の性質が大幅に変化した場合、著しく画像品質が低下する
- フレームレートが符号化後に決定されるため
 - 情報発生の大きさがフレームレートに反映されにくい⇒発生情報量一定、品質変動型の制御
 - 最大のフレーム間隔を規定できないため、リアルタイム伝送においてフレーム間隔の不一致が生じる (図2)

2.3 提案方式の制御系

提案方式の制御系を図3に示す。本方式では、

1. 空間的情報量制御
2. 時間的情報量制御
3. 統計的情報量制御

の各制御が仮想バッファメモリを仲介として、効果的な情報圧縮を行う。実際のシステムでは、空間的、時間的情報量制御は、それぞれ、符号化前処理における動き検出背景分離制御、フレームレート制御という形で実現し、統計的情報量制御は符号化ループ内での量

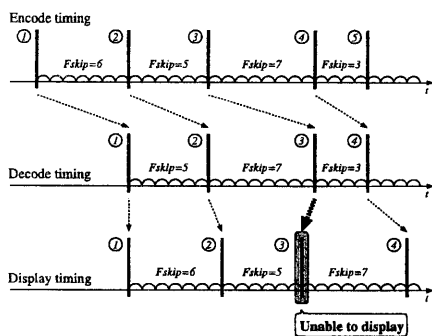


図 2: リアルタイム伝送におけるフレーム間隔の不一致

表 1: マクロブロックの分類

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

子化制御で行われる。本方式では、符号化前における画像性質の把握と、符号化中(後)の仮想バッファメモリからの帰還情報とから得られる最適動作点を、画像の性質により適応的に制御することを最大の特徴としている。以下に各制御の詳細について述べる。

2.3.1 空間的情報量制御

空間的情報量制御は、符号化前処理における動き検出背景分離により行われる。マクロブロック毎に、符号化雑音のない原画像同士で動き検出を行い、動き予測誤差を算出する。このとき、動ベクトルをその大きさが徐々に大きくなる順に検出し、動き予測誤差の絶対値和 D が閾値 T_H を下回った時点で、背景(非有意)ブロックと判定し、一切の量子化を行わず、動ベクトルのみを伝送することとする。さらに、動ベクトルが $(0, 0)$ の場合は、静止ブロックとして取り扱い、そのマクロブロックはスキップされる。

背景分離による空間的情報量制御は、不必要な情報発生を抑えるばかりでなく、ここでの有意ブロック数 S_{NS} から、フレームの持つ全体の情報量を予測し、次節で述べる時間的情報量制御や、統計的情報量制御に用いられる。このため、有意ブロックの数と情報発生との関係がある程度線形な関係を持つように、閾値 T_H を設定する。図 4 に、量子化ステップサイズを固定したときの、閾値 T_H と、発生情報量との関係を示す。

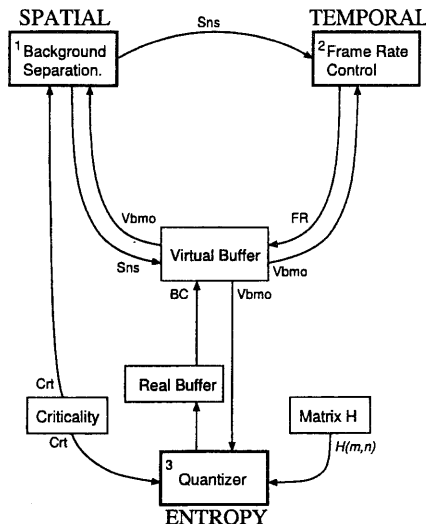


図 3: 提案方式の制御系

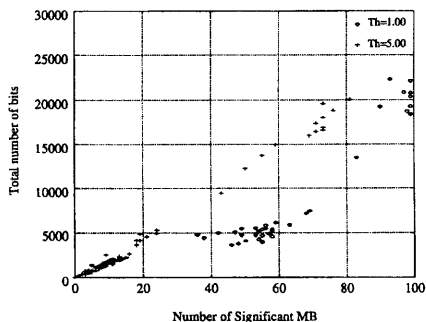


図 4: 閾値と線形性との関係例

この図より、閾値の違いにより線形性が大きく異なることが明らかである。

さらに、仮想バッファメモリの占有量を利用し、情報発生の大いシーンでは、より背景領域を増やすよう閾値 T_H を制御し、動物体により多くのビットを割り当てる。また、情報発生の小いシーンでは、通常では背景と分類されるような領域も、有意ブロックと判定し、さらに細かい量子化ステップでの背景部分の更新が行われる。

2.3.2 時間的情報量制御

時間的情報量制御は、符号化前処理部において、フレーム中に存在する有意ブロックの数、及び仮想バ

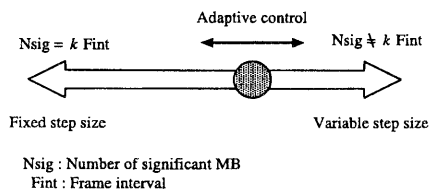


図 5: 時間的情報量制御と動作点

ファメモリの占有量により行われる。

先に述べた通り、有意ブロックの数が発生情報量を反映するよう空間的情報量制御がされている。このため、有意ブロック数とフレーム間隔とを比例関係に設定した場合、有意ブロック当たり割り当てられる情報量が一定となるため、量子化ステップサイズは、実際の発生量と予測値との偏差の吸収のために最小限の変動をする、すなわち固定ステップサイズ量子化に近い動作点となる。一方、フレームレートの変動幅を小さくすると、有意ブロック数にかかわらず、フレーム当たりの情報量が一定となるため、従来の可変ステップサイズ量子化に近い動作点となる。

本方式では、この動作点を、フレームの状態を考慮しながら、適応的に変化させることを特徴とする(図5)。

2.3.3 統計的情報量制御

本方式第3の制御として、統計的情報量制御を行う。これは符号化ループ内の量子化制御及び可変長符号化部から構成される。量子化制御は、画像の局所的絵柄(クリティカルティ)、人間の周波数方向での視覚感度特性 $H(m, n)$ 、後述の仮想バッファメモリ占有割合 V_{bmo} をパラメータとし、ステップサイズ QP が決定される。

$$QP = H(m, n) \cdot (aV_{bmo}^2 + bV_{bmo} + cC_{rt}^2 + dC_{rt} + eV_{bmo}C_{rt} + f) \quad (1)$$

定数 $a \sim f$ は、3 制御により所望の動作点となるよう決定される。

2.3.4 仮想バッファメモリ

実際の発生情報量を上述の各制御に反映させるため、仮想バッファメモリによる帰還を行う(図3中央)。空間的情報量制御による有意ブロック数 S_{NS} 、時間的情報量制御によるフレームレート FR 、統計的情報量制御の出力である実際のバッファメモリ蓄積量 $BC(t)$ とから、仮想バッファメモリ占有割合 $V_{bmo}(t)$ を出力する。フレームで発生した情報量を $BS(t)$ とすると、

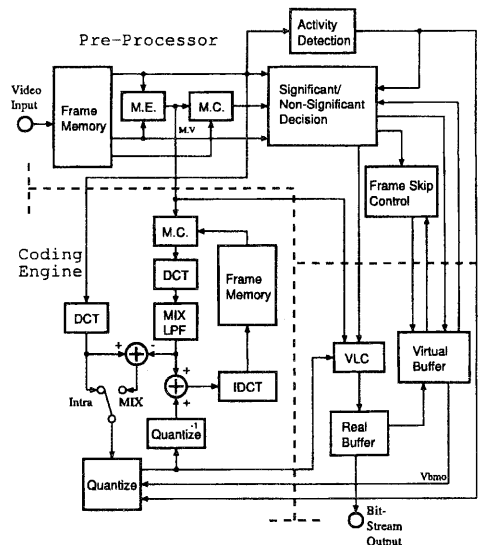


図 6: システムの全体構成

$$BS(t) = BC(t) - BC(0) + BR \cdot t \quad (2)$$

$$V_{bmo}(t) = V_{bmo}(0)$$

$$+ \frac{BS(t) - FR \cdot \frac{BR}{30} \cdot \frac{n}{S_{NS}}}{F_{max} \frac{BR}{30}} \quad (3)$$

となる。

3 超低ビットレートコーデックシステム

3.1 システムの全体構成

本方式のハードウェア上での実現を考慮したシステム構成を図6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。前述の空間的情報量制御、時間的情報量制御は共に前処理部で、統計的情報量制御は符号化部で、仮想バッファメモリ制御は伝送部で、それぞれ行われる。

まず、フレームメモリに蓄積された入力画像は、符号化雑音のない原画像同士で、動き検出が行われる。動き補償誤差は背景分離部に入力され、マクロブロック毎に有意・非有意(または静止)の判定を行う(空間的情報量制御)。判定閾値 T_H は、2.3.1での特性となるよう仮想バッファメモリの状態を反映させる。次に有意ブロック数 S_{NS} 及び仮想バッファメモリ占有割合によりフレームレート制御部(時間的情報量制御)で、フレーム間隔が決定される(2.3.2)。

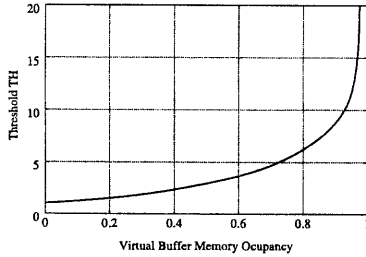


図 7: 有意ブロック判定閾値

このように符号化前処理部でフレームの持つ画像の特徴を把握し、それに応じて有意領域に割り当てる情報量を予め決定し、符号化部で実際の量子化を行う。

符号化部では、割り当てられた情報量を基に適応量子化を行う。量子化器は、視覚特性として、4種類の局所的絵柄変化度（クリティカリティ：Detail, Edge, Gradient, Flat）、周波数方向視覚感度の各パラメータが考慮される。また、割り当てられた情報量と実際の発生量とを一致させるために、仮想バッファの占有割合を帰還させる。

伝送部では、符号化部で発生した情報を実バッファに蓄積し、ビットストリーム出力すると同時に、前処理部での空間的情報量制御（有意ブロック数）、時間的情報量制御（フレームレート）と実バッファ蓄積量から、仮想バッファメモリを制御し、各制御に帰還する。

3.2 動作点の設定

本方式では、前処理部で把握したフレームの特徴から、図 5 に示す動作点を適応的に変化させ、有意領域を最大限に効率良く符号化することを特徴としている。

この動作点の設定が、本方式の符号化性能を大きく左右するところであり、各アプリケーションの空間的画像品質要求、時間的画像品質要求に応じて設定する。

双方向通信を意識するため最大遅延を規定し、かつ画像品質を優先した場合のパラメータ例を図 7, 8 に示す。この例では、情報発生が少ないシーンでは、バッファ占有状態が低くなるにつれ、背景部分が高品質に更新されるよう有意ブロック判定閾値 T_H を小さくする。また、予想発生量とフレーム間隔とが比例関係になるようフレームレート制御し、符号化部では固定ステップ量子化に近い動作点となる。一方、情報発生が多いシーンでは、規定したフレーム遅延（フレームスキップ）に抑えるため、バッファ状態が高くなるにつれ、可変ステップ量子化に依存する動作点としている。この場合、視覚的に破綻を起こさない範囲で、背景と判定されるブロックを増やし、不必要な情報発生を防ぐ。

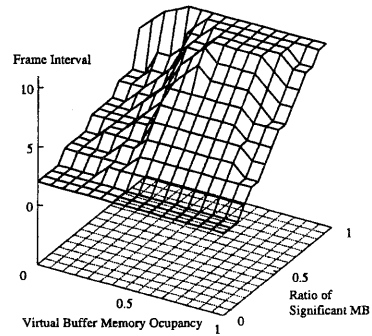


図 8: フレーム間隔パラメータ

表 2: 入力信号

画像形式	Y, C _r , C _b
画像サイズ	4:2:0 QCIF
フレームレート	30 枚/秒

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, 3 に示す。

4.2 実験結果

表 3: コーデックパラメータ

フレーム間隔	$2 \leq Fint \leq 11$
有意 MB 判定閾値	$1.0 \leq T_H \leq 8.0$
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)
符号化レート	BR=50.878kbit/sec
実バッファ容量	23.7kbits
エラー訂正	RS(63,59)
冗長度	9.253%
伝送レート	64kbit/sec

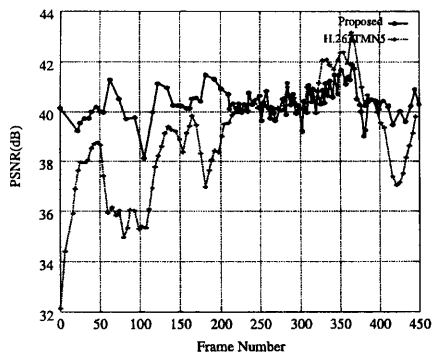


図 9: 各フレームにおける PSNR (SUSIE)

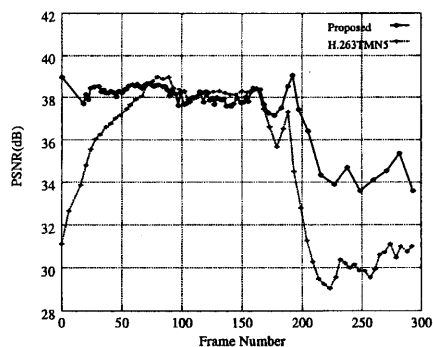


図 10: 各フレームにおける PSNR (RACE)

各フレームにおける PSNR を図 10, 9 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、各プロット間隔は処理フレーム間隔を表している。

画像 SUSIE は、ITU-R 標準画像で、4:2:0 QCIF に変換したものを用いている。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、後半 (約 190 フレーム以降) で、一台の車を追いかけるようにカメラがパニング動作をし、各フレーム毎に画像全体が大きく変化するものである。

まず、両画像共に、シーケンス前半 (約 50 フレームまで) において、提案方式の PSNR が従来方式を大きく上回っている。これは、提案方式では、最初の I-Picture より多くのビットを割り当て品質良く符号化している。その後の P-Picture では、背景領域としてそれを引き継いでいるため、動領域のみを効率良く符号化できることによる。視覚的にも動物体周囲に発

生するモスキート雑音明らかに低下している。

一方、画像 SUSIE の 50~150 フレーム、画像 RACE の 150 フレーム以降では、情報発生の多いシーンとなっており、ここでも提案方式が良い結果となっている。提案方式の前処理部における空間的、時間的情報量制御により、最適な空間間引きとフレーム間隔が与えられたためである。この部分で、提案方式がやや従来方式よりフレーム間隔が大きくなっているが、これは適応制御の結果であり、また視覚的に問題になる程ではない。

また、一部のシーンにおいて、従来方式の PSNR が提案方式を上回っている部分が見受けられる。しかし、主観評価においては、ブロック歪み、モスキート雑音、解像度等全てにおいて提案方式が上回っている。これは、提案方式の視覚特性による重み付け等は、必ずしも SNR の点では良い値にならないことが考えられる。

5 まとめ

超低ビットレート動画像符号化方式に関して、空間的情報量制御、時間的情報量制御、統計的情報量制御から成る新たな符号化方式を提案した。

符号化前処理部において、空間的情報量制御 (背景分離制御)、時間的情報量制御 (フレームレート制御) を行うことで、前もって画像の持つ特徴を把握し、それに応じてフレーム中の有意領域に割り当てられる情報量を予測する。符号化ループ内では、統計的情報量制御 (適応量子化制御) により、割り当てられた情報量を最大限に効率良く符号化し、出力する。これにより、画像の性質に応じて、システムの動作ポイントを適応的に制御することが可能となり、大幅な画像品質の向上を得た。

今後は、詳細な情報発生予測を備えた符号化方式について検討する予定である。

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: "ITU-T Recommendation H.263; Video coding for low bit rate communication", March 1995
- [2] ITU-T Study Group 15: "VIDEO CODEC TEST MODEL, TMN5", January 1995
- [3] 宮地, 松本: "情報量予測型適応可変フレームレート制御方式", 情処研報, 96-AVM-15, Vol.96, No.120, pp.27-34 (1996)