

## 背景分離型適応可変フレームレート動画像符号化方式

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 以下の超低ビットレートで動画像を符号化することを目的として、背景分離手段および適応可変フレームレート制御手段を有した動画像符号化方式についての検討を行う。映像信号から背景を分離し、さらに動領域においても有意領域を効率的に抽出し符号化する手法を提案する。また、画像の持つ情報量や画面の動き量を利用して、発生符号量の大幅な削減をし、なおかつ視覚的に自然な動画像となるような可変フレームレート制御方法を提案する。また、これらの提案方式に加え、とくに低レートにおいて効率良く符号化を行うためのいくつかの手法を示す。

## Variable Frame Rate Image Coding with Background Separation

Satoshi MIYAJI      Shuichi MATSUMOTO

KDD R&D Laboratories

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

This paper describes a very low bit-rate (less than 64kbps) image coding scheme with background separation and adaptive frame rate control. A method of separating background from image and coding significant blocks extracted efficiently from moving area is suggested. And a method of variable frame rate control exploiting the number of significant macro block and length of motion vector is proposed. In addition to these proposed method, some techniques are shown to achieve advanced very low bit-rate image coding.

## 1 はじめに

ISDN 回線や移動体における動画像伝送を目的として、64kbps 程度の超低ビットレート画像符号化の研究が活発化している。従来の研究の流れとして、高度な動き補償の導入等いわゆる MC-DCT の波形符号化を踏襲した技術と、画像の中身に立ち入った俗に言う知的符号化とがある。

本稿で新たに提案する符号化方式は前者に属する。そして、その要点は

1. 限られた符号化ビットレートの下で、符号化に必要な最小限の領域をいかにうまく抽出するか。
2. 抽出された領域をいかにひずみを少なく符号化するか。

の2点にある。このため、まず1に対しては、符号化に先立つ前処理として背景と動物体とを分離し、さらに動物体に対しては有意・非有意判定を行い、有意領域を特定する。2に対しては、有意領域をその面積に比例したフレームレートで符号化する。その際、フレームレートは H.263 とは異なり、各フレームの符号化に先立ち決定されるため、各有意領域に合わせた無駄のない符号化が行えるばかりでなく、フレーム間引きによる動きの不自然さは、画面の動き量をフレーム間引き制御に組み込むことにより、最小限のものとなっている。

また、符号化自体は、従来の MC-DCT に基づくが、筆者らが先に 45Mbps HDTV 符号化用に開発した MIX モード符号化、視覚特性を反映させた量子化制御、の2つの技術が導入され、より効果的な符号化が行われている。

本稿では、これらの主要技術ならびにシステム構成について述べる。

## 2 システムの全体構成

図1に符号化器側の全体構成を示す。符号化の対象画像は CIF もしくは QCIF を想定している。符号化器は、前処理部、符号化部、伝送部から構成されている。前処理部では、後述する、1. 動き検出背景分離と有意領域の抽出、2. 画像の局所的絵柄の変化度(クリティカリティ)のマクロブロック毎の判定、3. 符号化モード判定、4. フレームレート制御(フレームスキップ数の決定)の4機能を実行し、符号化部を制御する。

符号化部では、後述するミックスモード符号化と、視覚特性反映量子化を特徴とする動き補償フレーム間 DCT 符号化を行い、画像圧縮を行う。伝送部は、フレーム多重ならびにバッファメモリから構成され、バッファメモリ占有割合に応じて符号化部ならびに前処理部を適応制御する。なお、このバッファメモリは、可変のフレームレートを吸収するため、最大フレームスキップ数に相当する伝送ビット数を蓄積できる容量を有する。

伝送部から出力された画像データは、7.2kbps 音声信号(CELP)を多重され、誤り訂正 RS(239,255)を施され、所定の伝送フォーマットで送出される。

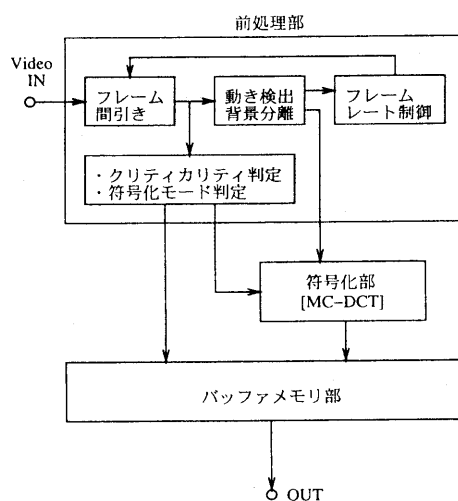


図1: 符号化装置の構成

## 3 各部の機能

### 3.1 前処理部

図2に前処理部のブロック図を示す。以下に各機能を説明する。

#### 3.1.1 動き検出

各マクロブロック毎に前処理部のフレームメモリに蓄積されている直前の処理対象フレーム(フレームメモリには、後述する有意マクロブロックと判定された領域のみ初期フレームに対して内容が更新された背景画

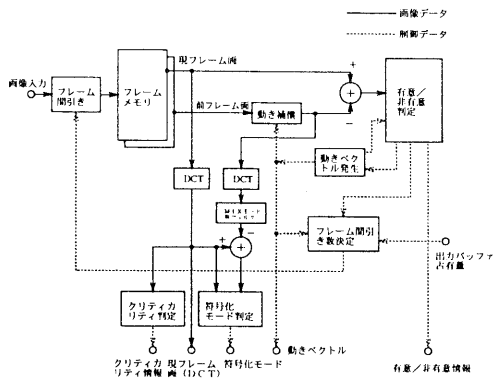


図 2: 前処理部のブロック図

面が蓄積されている)と入力フレームとの間で動き検出が行われる。原則的に直前の処理対象フレームには符号化雑音が重畳されていないため、符号化再生フレームを用いるより正確な動き検出を行うことができる。

### 3.1.2 背景分離と有意マクロブロックの抽出

初期フレーム(リフレッシュフレーム:全て Intra で符号化される)はフレーム間予測の起点となる背景画面となるため、符号化部ではフレームスキップ数を比較的大きく設定し高品質に符号化される。この初期非符号化フレームが背景フレームメモリに入力され、その後に続く符号化フレームに対して背景を分離し、マクロブロック単位に有意・非有意判定を行うことにより入力フレームの有意マクロブロックが抽出される。

これにより、非有意ブロックは一切の予測誤差値の符号化をせず、動きベクトル無し(背景)か有り(動領域)のみで符号化は完了し、不必要な情報発生をさけるとともに厳密には正しくないが、視覚的に高品質の画像が再現できる。

マクロブロック単位の有意・非有意判定は、背景フレームを更新するかどうかを決定するため、特にきめ細かく行う必要がある。

そこで、マクロブロックを図 3 に示す 4 つの(8 画素×8 ライン)ブロックに分割し、各ブロックの動き補償誤差を式 1 で求め、その最大値が所定の閾値  $T_h$  より小さい場合、非有意マクロブロックと判定する方法を採用した。これにより局所的に有意な情報がある程度救済され、背景メモリの更新をより正確に行うこと

が可能となる。

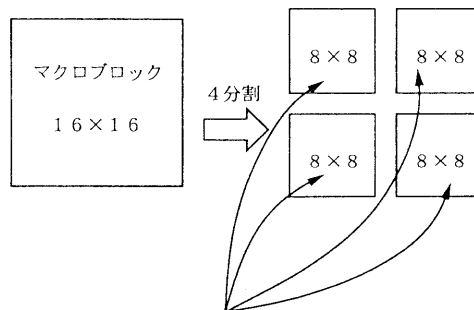


図 3: 差分絶対値

$\max(S1, S2, S3, S4) < T_h \rightarrow$  非有意マクロブロック  
 $\max(S1, S2, S3, S4) \geq T_h \rightarrow$  有意マクロブロック

$$D = \max_n \left( \frac{\sum_{i=0}^{S3} |An_i - Bn_i|}{8 \times 8} \right) \quad (1)$$

$(n = 0, \dots, 3)$

動きベクトルは(0,0)から順にベクトル長が大きくなる順番に検出され、前述した非有意マクロブロックが検出された時点のベクトルを動きベクトルとする。

非有意マクロブロックと判定されなかったブロックに対しては、動き補償誤差のブロック最大値が最小となる動きベクトルを選択し、同時にフレームメモリの内容を、次の処理対象フレームの背景分離に用いるため更新する。

背景分離動作のフローチャートを図 4 に示す。

なお、マクロブロックの有意/非有意/静止の取り扱いについて、表 1 にまとめる。

表 1: マクロブロックの分類

分類	ID	判定条件	符号化
静止	0	$D < T_h$ , 動ベクトル 0	—
非有意	10	$D < T_h$	動ベクトル
有意	11	$D \geq T_h$	MC-DCT

また、有意・非有意判定閾値に関しては、動きベクトルを可変長符号化する際、符号割り当てはベクトル

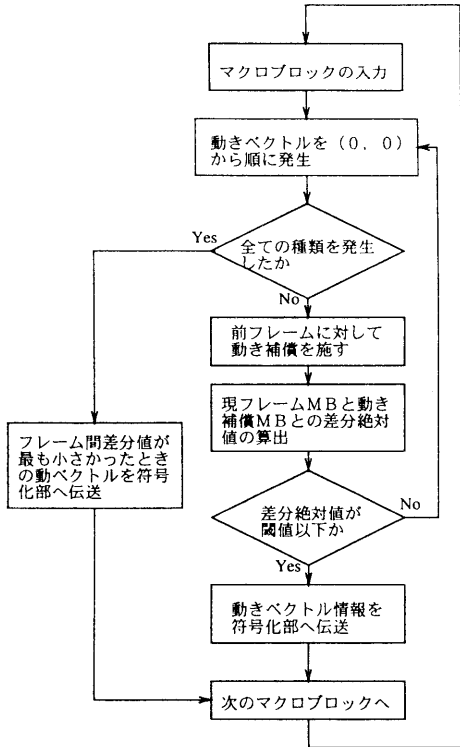


図 4: 背景分離処理のフローチャート

の成分の大きいものにより多くのビットを割り当てるため、動きベクトル情報しか持たない非有意マクロブロックの場合、非有意と判定される範囲内で、動きベクトルは、そのスカラ長ができるだけ小さくなるように選ばれるようにすることが望ましい。

従って、その判定閾値は、図5に示すような動きベクトルのスカラ長を指数に持つ単調減少関数とする。

### 3.1.3 クリティカリティ検出

画像の符号化雑音は、原画像の変化の度合より検知の程度が異なることが、経験的に知られている。すなわち、より変化している部分ほど、重畳した雑音を検知されにくく、いわゆる雑音のマスキング効果が存在する。従って、DCT係数の量子化を行う場合、雑音の目に付きにくい変化部では粗く量子化し、その分情報を節約して雑音の目に付きやすい平坦部へとビットをより多く割り当てて量子化することにより、再生画

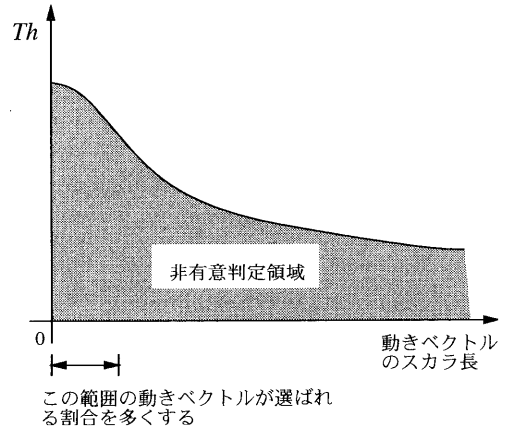


図 5: 判定閾値  $Th$

像品質のばらつきを少なくすることができる。

本システムでは、各マクロブロックにおける変化の度合(クリティカリティ)を次の方法により求める。有意と判定されたマクロブロックに対して、そのDCT変換係数と、DCT係数毎にあらかじめ設定した閾値とを比較し、この閾値を超えた係数をカウントする。カウントされた有意な係数の個数に応じて、4種類(Detail, Edge, Gradient, Flat)に分類する。

1. 符号化対象マクロブロック内のDCTブロック(8画素×8ライン)の各係数に対して、クリティカリティ判定閾値を上回る係数を有意係数としてカウントする。
2. 有意係数を4ブロック分(1マクロブロック分)累計する。
3. 表2により4種類に分類する。

表 2: クリティカリティ判定表

有意係数の個数	クリティカリティ
0 ~ $n_0$	Flat
$n_0 + 1$ ~ $n_1$	Gradient
$n_1 + 1$ ~ $n_2$	Edge
$n_2 + 1$ ~ 256	Detail

ここで分類されたクリティカリティ情報は、符号化部へと入力され、量子化制御に用いられる。

### 3.1.4 符号化モード判定

通常符号化モード判定は、符号化部で行われるが、本方式では原画像レベルでモード判定を行う方式を採用した。これは、情報量最小の状態での最適なモードが必ずしも量子化に最適ではないということに基づいている。

符号化モードは、後述する Inter と Intra を DCT 係数上で混在させた MIX モードと Intra モードの 2 つが用意されている。

ミックスモードは、帰還する量子化雑音の影響を受けるフレーム間予測と、フレーム内高次 DCT 係数とでは、発生する雑音がフレーム内 (Intra) の方が視覚的に受け入れやすいという特性に基づいている。

その動作概要を図 6 に示す。動き補償を施し DCT 変換後のブロックに対し、低域部が“1”、高域部が“0”からなるマトリックスにより高域部分の DCT 係数を切り捨てるフィルタ処理を行う。次に DCT 変換後の符号化対象画像からの差分を取る事により、低次係数が Inter、高次係数が Intra となる MIX モード DCT 係数が得られる。

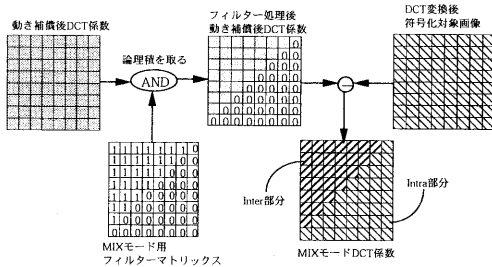


図 6: MIX モード処理概要

符号化対象画像の Intra モードの DCT 係数と MIX モードの DCT 係数との間で、各 DCT 係数の大小比較を行い小さい係数の多い方のモードを選択する。

ただし、 $C_r, C_b$  信号の場合、MIX モードは用いず、 $Y$  信号で判定されたモードに合わせて以下のように選択される。

- $Y$  信号が Intra モードの時 →  $C_r, C_b$  信号は Intra モード
- $Y$  信号が MIX モードの時 →  $C_r, C_b$  信号は Inter モード

### 3.1.5 フレームレート制御

従来の低レート可変フレーム符号化の標準的な方式である H.263 の TM5 では、前フレームの符号化結果から目標符号化ビット数を定め、量子化ステップサイズを決定し、現フレームの符号化を行う。その発生情報量の目標符号化ビット数からの偏差を吸収するように次フレームまでの間引き数を変化させ、フレームレートの制御を行っている。

しかし、この方法は、フレームレートの制御は現フレームを符号化した後の副産物的な要素にすぎず、画像の持つ特徴が十分に反映されていない。例えば、画像の持つ情報量が大きい場合でも、従来の手法ではそれがまず量子化ステップに反映されてしまい、画像の品質が著しく低下してしまう。フレームレートは、この量子化ステップで吸収しきれない情報量の変化するだけである。また、動きの大きな物体が存在するような場合でも、その情報がフレームレート制御に反映されていないため、非常に動きの不自然な画像になってしまう。

そこで、現フレームの符号化を行う前に、画像の持つ情報量や動き量を反映させてフレーム間引き数を決定する新たなフレームレート制御方式を提案する。

本提案方式は、現フレームの符号化を行う前に、次の 3 つのパラメータ

- 現フレームの有意マクロブロック割合 (予想発生情報量)
- 現フレームの動きベクトルのスカラ長
- 出力バッファ占有状態

から、次フレームまでのフレーム間引き数を決定するものである。

前もって、フレーム間引き数を決定することで、動き情報を利用することができるだけでなく、本符号化方式では符号化前に背景の分離を行っているので、予想される発生情報量からフレームレートを制御すれば、1 つの有意マクロブロックあたりに割り当てられる情報量をできるだけ一定にすることができ、量子化器の制御を効率良く行うことができる。

有意マクロブロックの割合を  $S_{mb} (0 \leq S_{mb} \leq 1)$ 、現フレーム中の動きベクトルのスカラ長の最大値を  $L_{mv} (0 \leq L_{mv} \leq 16\sqrt{2})$ 、出力バッファの占有割合

を  $BMO(0 \leq BMO \leq 1)$  とすると、フレーム間引き数を  $F_{skip}$  は、

$$F_{skip} = \left( \alpha S_{mb} + \beta - \frac{L_{mv}}{\gamma} \right) \left( 1 + \frac{BMO - 0.5}{\delta} \right) \quad (2)$$

( $\alpha, \beta, \gamma, \delta$ は定数)

で決定する。

### 3.2 符号化部

符号化部では、前処理を施された画像データ(有意マクロブロック)に対して、動き補償、DCTおよび量子化を行う。

図7にブロック図を示す。

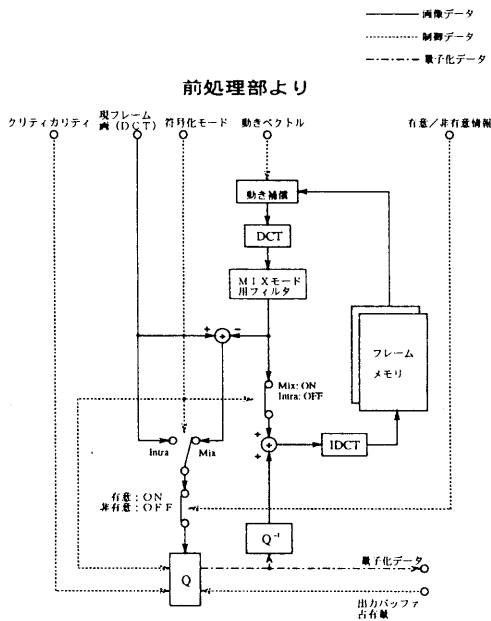


図7: 符号化部のブロック図

#### 3.2.1 5パラメータによる適応量子化制御

本符号化方式では、以下の5パラメータによる適応量子化制御を行う。

- クリティシティ
- 係数位置
- 符号化モード (Intra/MIX)

- $Y$  or  $C_r, C_b$

- バッファメモリ占有割合 (BMO)

クリティシティによる量子化制御では、画像の局部的絵柄の変化度合(クリティシティ)が小さい平坦部分はより細かい量子化を行い、逆に変化の大きい部分に対しては多少粗い量子化を行う [1]。

係数位置での量子化制御では、人間の視覚特性は周波数の低い領域に対して雑音を感知しやすいことから、DCT係数の位置に応じて重み付けを行う [2]。

また、バッファメモリ占有割合を量子化ステップに反映させ、3.1.5で述べたフレームレート制御では吸収しきれない情報発生量の変動をできるだけ小さくするように作用させる。

### 3.3 バッファメモリ部

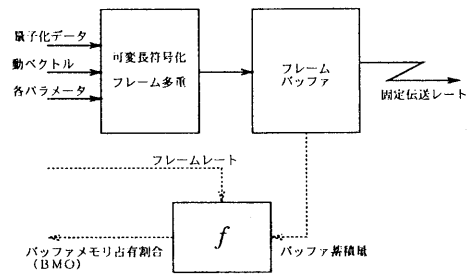


図8: バッファメモリ部の構成

バッファメモリ部の構成を図8に示す。バッファメモリ部では、前処理部および符号化部で発生した各情報を可変長符号化し、フレーム多重を行ってビットストリームを生成する。

また生成したストリームを出力バッファに蓄積し、一定レートで出力する。

#### 3.3.1 可変長符号化とフレーム多重

本符号化方式におけるビットストリームの構成を図9に示す。

本方式で注目すべき点は、マクロブロックレイヤーにおけるMBTYPEである。これは、3.1.2章の表1に示した通り、マクロブロックが静止の場合1bit("0")となり、このブロックに関するその後に続く情報は一切伝送しない。マクロブロックが非有意の場合2bit("10")

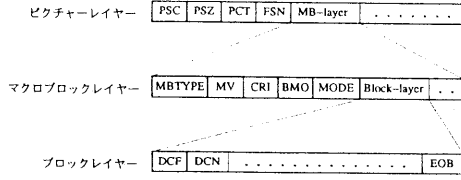


図 9: ビットストリーム構成

となり、次に続く情報として MV のみを伝送する。有意ブロックの場合には 2bit("11") となり、後に続く情報すべてを伝送する。

### 3.3.2 バッファメモリ占有割合

本方式では、フレームレート制御や量子化制御にバッファメモリ占有割合 ( $B_{MO}$ ) を用いている。これは、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の割合であるが、可変フレームレートの場合、フレーム間引き数が 0 枚の場合と 29 枚の場合とでは、次フレーム情報発生までの時間が異なるため、例えば、実際のバッファメモリ内の情報蓄積量が同値の場合でも、間引き数 0 枚の場合に比べて 29 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。

各フレームレートに対応したバッファ制御を行うために、図 10(a) 左側のバッファの占有割合を  $B_{MO}$  として用いる。

図 10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、 $\alpha(t)$  を発生情報量、 $BR$  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 10(a) は可変フレームレートに対応した理想的な 2 バッファ構成で、左側のバッファの占有割合  $B_{MO}(t)$  を次式で定義する。

$$B_{MO}(t) = \frac{B(0) + \alpha(t) - \frac{BR}{FR} \cdot \frac{S_t(t)}{N_s}}{B(0) + \frac{BR}{FR}} \quad (3)$$

上式より、

$$B_{MO}(0) = \frac{B(0)}{B(0) + \frac{BR}{FR}} \quad (4)$$

したがって、

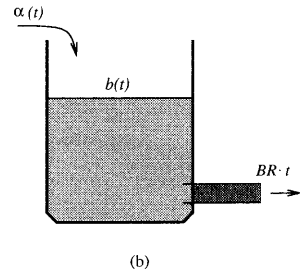
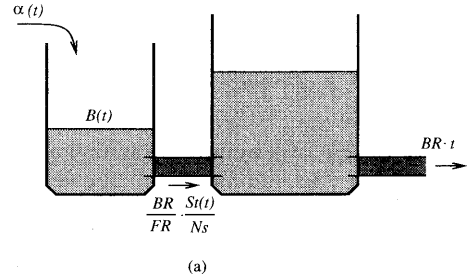


図 10: 出力バッファ概念図

$$B(0) = \frac{\frac{BR}{FR} \cdot B_{MO}(0)}{1 - B_{MO}(0)} \quad (5)$$

また、図 10(b) が本方式の実際のバッファであり、その蓄積量は

$$b(t) = b(0) + \alpha(t) - BR \cdot t \quad (6)$$

であるので、式 (3)、式 (5) および式 (6) より、実際のバッファ蓄積量  $b(t)$  を次の式で補正して  $B_{MO}(t)$  を求める。

$$B_{MO}(t) = \frac{b(t) - b(0) + BR \cdot t - \frac{BR}{FR} \cdot \frac{S_t(t)}{N_s}}{\frac{\frac{BR}{FR} \cdot B_{MO}(0)}{1 - B_{MO}(0)} + \frac{BR}{FR} + B_{MO}(0)} \quad (7)$$

ただし、

$$FR = \frac{30}{1 + F_{skip}}$$

となる。

## 4 シミュレーション実験

本提案方式の有効性を確認するために、シミュレーション実験を行った。QCIF 画像 5 種類を用いて、本方式と標準方式 (H.263, TM5) との比較を行った。

実験の条件を表 3,4 に示す。

表 3: 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
( $C_r, C_b$ )	88 画素 × 72 ライン
フレームレート	30 枚/秒
画像数	5 種類

表 4: 前処理, 符号化

フレーム間引き数	$0 \leq F_{skip} \leq 29$
有意 MB 判定閾値	$1.0 \leq Th \leq 5.0$
動きベクトル発生	±16 画素, 1/2 ステップ
クリティカリティ	Flat, Gradient, Edge, Detail
符号化モード判定	Intra, MIX (C 信号は Inter)
量子化制御	5 パラメータ適応量子化

シミュレーション実験の結果、提案方式の視覚的品質が大幅に従来方式を上回っている。比較的動き予測的中している領域 (平行移動等) では、背景分離 (非有意判定) の作用により量子化を一切行っていないため、画素値は必ずしも正確ではないものの、符号化雑音が存在しないため、視覚的に非常に質の高い画像が得られている。また、動きの複雑な領域については、背景分離と適応フレームレート制御により、品質を維持するための一定の情報量割り当てが保証されているので、従来方式に比べ、明らかに符号化雑音が少なくなっている。

また、フレームレート制御に動きベクトルの大きさを考慮しているので、大きな動きのある画像に対しては、多少量子化ステップが粗くなるものの、動きの滑らかな画像を得ることができる。

## 5 まとめ

背景分離、適応フレームレート制御および適応量子化制御を有する超低レート動画画像符号化装置に関する検討を行った。

背景分離により、背景領域 (非有意領域) については符号化を行わないため量子化雑音が発生せず、動領域 (有意領域) にはより多くの符号化ビット数を割り当てることが可能となり、画像品質の向上が可能となった。

また、1 フレームを占める動領域の割合、動領域の動き量、出力バッファの状態からフレーム間引き数を決定することで、各フレームの画像品質をできるだけ一定に保ちつつ、かつ視覚的に動きの自然なフレームレート制御を行うことができた。

さらに、5 パラメータによる量子化制御および MIX モードにより、低レート時においても効率良く MC-DCT 符号化を行うことが可能となった。

今後は、背景分離の閾値や量子化制御関数の最適化を行う予定である。

## 謝辞

日頃御指導頂く国際電信電話 (株) 研究所の浦野所長、村上次長、古賀画像通信グループリーダーの各位に感謝いたします。

## 参考文献

- [1] 浜田, 松本: “画像の局所的変化度による雑音マスキング効果を考慮した直交変換係数の最適量子化法”, 信学論, **J75-B-I**, 12, pp.791-801(1992)
- [2] 浜田, 松本, 村上: “量子化器最適化に基づく動き補償 DCT 方式の特性限界”, 信学論, **J78-B-I**, 11, pp.689-699(1995)
- [3] ITU-T Study Groups: “DRAFT RECOMMENDATION H.263; Video coding for narrow telecommunication channels at < 64kbit/s”, March 1995
- [4] ITU-T Study Groups: “VIDEO CODEC TEST MODEL, TMN5”, January 1995