

■2群（画像・音・言語）-5編（画像符号化）

8章 符号化制御

（執筆者：杉山賢二）

■2群 - 5編 - 8章

8-1 符号化制御とは

(執筆者：杉山賢二) [2009年12月 受領]

画像符号化は極簡単なものを除いて可変長符号化 (VLC) を用いるので、画像により発生符号量が大きく変化する。また、符号化規格は、新しいものほどエンコーダ側の自由度が大きくなっており、処理方法により大きな画質の差を生じる。したがって、与えられた符号量 (ビットレート) の条件に対して最適な符号化処理を行い、最も高い再生画質を得るために符号化制御が必要となる。そこでは、ビットレートの制御と最適符号化の二つの要素があるが、レートについての要求は伝送路 (媒体) に大きく依存し、最適符号化は処理量 (ハードウェア) の制約がある。なお、静止画像でも記録画像数を維持するために符号化制御は必要であるが、動画像符号化においてより高度なものが要求されるので、本章では主に動画像符号化に関して説明する。

動画像符号化とその制御の処理構成を図 8・1 に示す。レート制御は、バッファ (Buffer) に残留している情報量 (充足度) を観測し、量子化 (Q) を制御するのが基本である。発生符号量を抑えたい場合は、量子化ステップ幅を大きくし、粗い量子化を行う。他に被符号化画像自体にプリフィルタ (LPF) を施す場合や、低ビットレートではフレームレートや画像フォーマット (画素数) を制御する場合もある。これは、原画像に対する忠実度より、再生画像の画質バランスを優先する場合に行われる。

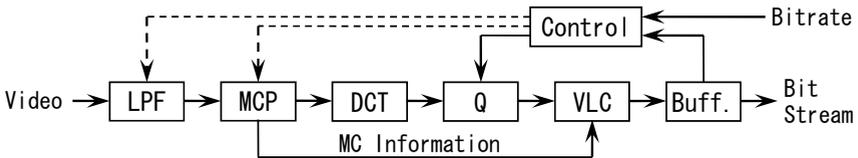


図 8・1 動画像符号化とその制御

一方、エンコーダで変更できる処理は表 8・1 に示されるように多岐にわたる。量子化ステップ幅がレート制御に使われる以外は、再生画質向上やレート低減のための処理モードの選択肢として設定されている。各処理モードで符号量と画質 (ひずみ量) を見積もり、最適なものを選択することで、所定レートで最も画質の良い符号化が実現できる。この最適化は、動き補償予測 (MCP) での処理が中心となる。

表 8・1 エンコーダで変更可能な処理

変更対象	変更部分	処理概要
画像	被符号化画像	ノイズ軽減、高周波成分抑圧
ピクチャ	タイプ (IPB)、フレーム/フィールド	シーン・動き程度で変更
動き補償予測	予測方向、ブロックサイズ、動きベクトル	ひずみとレートの関係で最適モード選択
量子化	ステップ幅、閾値、マトリクス	ビットレート・再生画質の制御

■2群 - 5編 - 8章

8-2 CBR と VBR

(執筆者：杉山賢二) [2009年12月 受領]

ビットレートの大局的な制御方法としては、固定転送レート (Constant Bit Rate) と可変転送レート (Variable Bit Rate) の2種類がある。VBR は理想的なレート配分が可能になり、平均レートに対して高画質化できるが、デジタル放送など無線システムでは周波数割当ての制約から CBR とならざるを得ない。

CBR でのレート制御は、バッファ充足度により行われるが、複数のピクチャタイプ (I, P, B-Picture) をもつ符号化は簡単でない。シーンに応じた I, P, B 間の符号量バランスを Group of Picture 単位で設定し、それと実符号量の差で制御する。

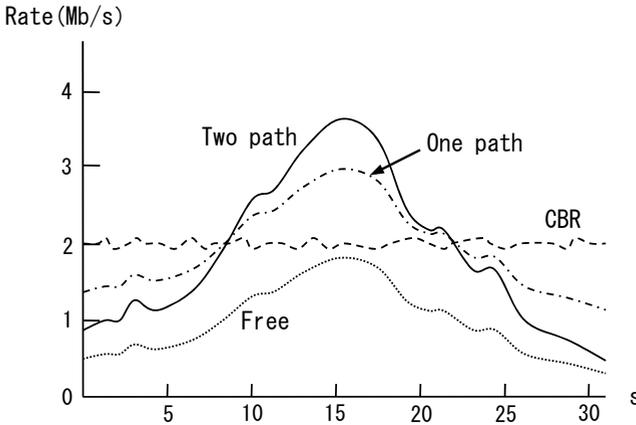


図 8・2 転送レートの様子

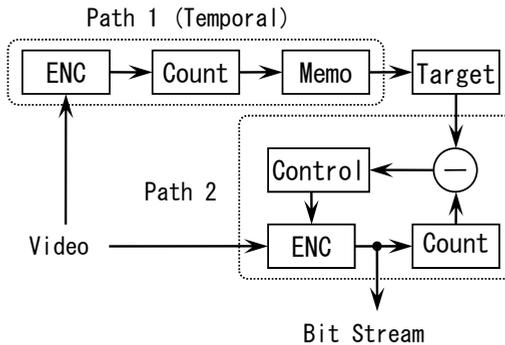


図 8・3 2パス符号化制御

VBR の制御方法は、用途 (伝送・記録媒体) により、フリー、2パス、1パスの3タイプに分けられる。CBR を含めたこれらのビットレートの様子を図 8・2 に示す。フリーは、総符号量を気にせずに画質が一定になるようにするもので、ダウンロード用のファイルなどに用

いられる。2パスは、蓄積媒体など非リアルタイムシステムで可能なもので、図 8・3 のようにフリー同様な仮符号化 (Path 1) で発生符号量を観測し、それに基づいて目標を設定して本符号化 (Path 2) を行う。これにより総符号量を一定にしながら最適なレート配分を行うことができる。1パスはリアルタイム制御で、大局的発生符号量に応じて制御するので、非常に大きなバッファの CBR 制御ともみなせる。なお、VBR であっても最高レートは制限する必要がある。

■2群 - 5編 - 8章

8-3 バッファ管理

(執筆: 杉山賢二) [2009年12月受領]

大局的な制御に対して、フレーム単位以下の制御ではバッファとの関係が重要となる。動画画像符号化では、VBR/CBRに関わらず図8・4に示されるようにエンコーダ側とデコーダ側の両方にバッファをもつ。バッファ残量(充足度)は、発生符号量により変化するが、デコーダ側はエンコーダ側の逆となるので、合計残量は常に一つのバッファ分となる。

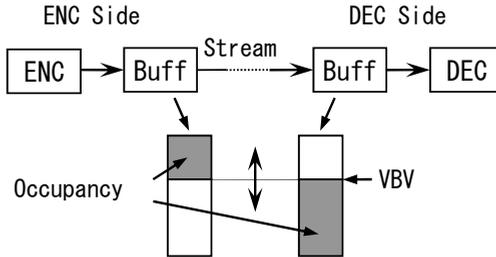


図8・4 バッファと充足度

符号化規格では、バッファ容量と充足度の推移に規定が必要である。具体的には復号化側のバッファにおいて、ピクチャ単位など所定タイミングでデータがデコーダに渡されるとし、その充足度を Video Buffer Verifier とする。VBV の変化の例を図8・5に示すが、CBR の場合にはバッファに入る量は一定なので、増加は一定で減少はピクチャ単位となる。

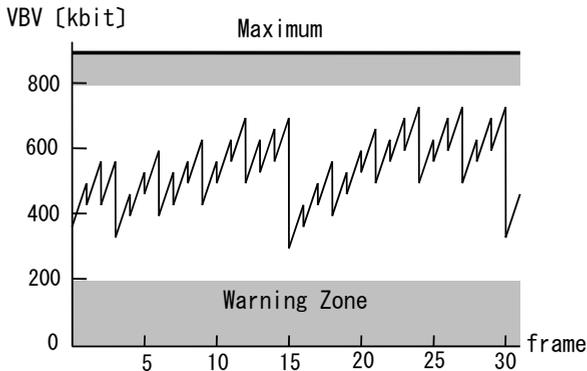


図8・5 VBVの様子

符号化の制御では、VBV が 0 から最大値(バッファ容量)の範囲に収まるようにする。VBV が 0 とは、エンコーダ側ではデータがバッファを溢れ、デコーダ側では復号化できるデータが無くなった状態である。VBV が 0 になるのを避けるため、発生符号量の増加を急激

に押さえ込むと画質が大きく劣化するので、**VBV**を高めに保持するのが無難である。逆に制御しきれず最大値になりそうな場合は、スタッフ（無効）ビットを挿入すればよい。なお、**VBR**の場合もレートが変化するだけで、制御の考え方は同様となる。

一方、バッファ容量が多いと制御は容易になるが、ハードウェアを用意する必要があると共に、遅延の問題を生じる。符号化遅延はバッファ容量をレートで除した値なので、レートに対してバッファ容量が少ないほど遅延は少ない。遅延を少なくしたい場合は、最大値まで使い切らずに使用範囲（実効容量）を少なめに設定する。

■2群 - 5編 - 8章

8-4 モード選択

(執筆者：杉山賢二) [2009年12月受領]

動画画像符号化では、特に動き補償予測において非常に多くの符号化モードが設定されている。これらはエンコーダ側で選択し、どのモードが選ばれたかは情報伝送される。処理モードは、表 8・1 で示したようにピクチャ単位での変更されるものもあるが、多くは表 8・2 に示されるマクロブロック (16×16 画素) 単位のものである。具体的内容は、規格 (Profile) やピクチャタイプで異なるが、H.264 において極めて多くなっている。

表 8・2 マクロブロック符号化モード

処理モード	概要
符号化/非符号化	予測残差が少ない場合、符号化しない方が良い
イントラ/インター	予測しないでフレーム内符号化とするか
方向 (参照画像)	順方向、逆方向、双方向、参照画像の選択
フィールド/フレーム	インターレース走査画像での予測単位及び DCT の並び
ダイレクトモード	動きベクトルを共用して、情報を削減するか
ブロックサイズ	細かいほど予測残差は減るが、動きベクトル増加
動きベクトル	動きベクトルの情報量も考慮した最適ベクトルの選択
マクロブロック量子化	量子化ステップ幅のマクロブロック単位での変更

■2群 - 5編 - 8章

8-5 最適化処理

(執筆者：杉山賢二) [2009年12月受領]

モード選択は、与えられた符号量に対して最適なものとしたい。そのためには、各マクロブロックにおいて、各モードで仮処理を行い、最も効率が良いモードを選択することになる。

画像全体での符号化効率は、図8・6のようにビットレート対PSNR (Peak Signal to Noise Ratio) で表される。効率が良いとは、符号量 R (Rate) に対してひずみ量 D (Distortion) が少ない (PSNR が高い) ということであるが、モードを変更すると R と D の両方が変化し、モード優劣の判断ができなくなる。そこで、 D を λ で R に換算し、 R と D を一元化したコスト関数 $C=R+\lambda D$ を得て、 C が最小になるものを選ぶ。しかし、 R と D の関係は図8・6のように線形でなく、画像により異なる。そこで、ラグランジュ未定係数の考え方をを用いることで、所定条件下における λ を得る。

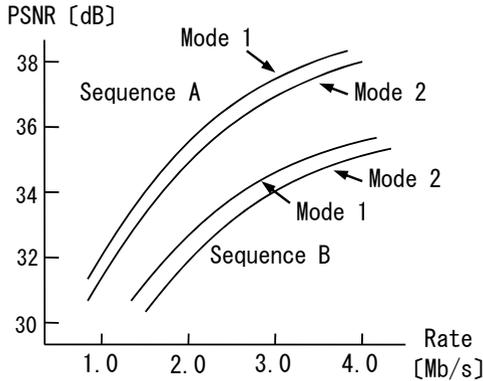


図8・6 ビットレートとPSNRの例

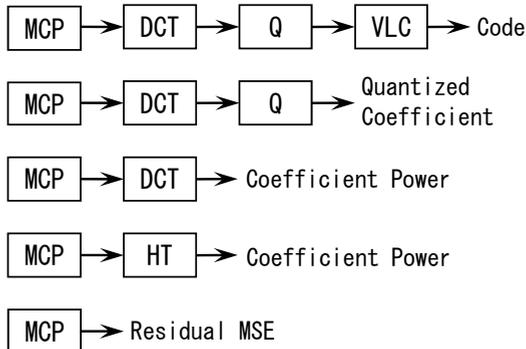


図8・7 モード判定方法

この処理では、各モードでの符号量 R も知る必要がある。すべての符号化処理を行い、発生符号量を計測すればよいが、モード数に比例して処理量が極めて多くなる。簡略化手法として、以前は予測残差の MSE 値 (Mean Square Error) で評価していたが、中間的なものが増えている。これらを図 8・7 に示すが、VLC のみ省き量子化された係数値で評価するもの、量子化も省き係数電力値で評価するもの、更に DCT を加減算だけで済むアダマール変換 (HT) にしたものなどがある。HT のみなら処理量はかなり少ない。なお、動きベクトルの符号量もあるが、算出は比較的容易である。また、ビットプレーン符号化など各マクロブロックで符号量が確定しない場合は、近似にならざるを得ない。

一方、ひずみ量 D は MSE が基本で、これは PSNR のもととなるものである。PSNR 評価ではこれで最適となるが、主観画質において最適とするためには、主観に対する考慮が必要となる。具体的には MSE に視覚特性の重み付けをするが、特徴的な劣化に関しては、その量的評価も必要となる。また、特定モードで主観評価が悪い場合は、そのモードの D を補正する。