

## [チュートリアル招待講演] MPEG ビデオ符号化方式

渡辺 裕<sup>†</sup>

<sup>†</sup> 早稲田大学大学院国際情報通信研究科

〒367-0035 埼玉県本庄市西富田 1011

E-mail: <sup>†</sup>hiroshi.watanabe@waseda.jp

**あらまし** MPEG は Moving Picture Experts Group の略称であり、ISO/IEC JTC1 におけるオーディオビジュアル符号化標準の開発機関及び規格の両者を意味する。MPEG ビデオ規格の代表的なものに MPEG-1, MPEG-2, MPEG-4, MPEG-4 AVC がある。このうち、MPEG-2 は DVD と HDTV に、MPEG-4 AVC は携帯端末用ワンセグ放送、BD 用映像符号化、YouTube などでも用いられている。本稿では、ネットワーク帯域の消費量が大きいと考えられる MPEG-2 のビデオ符号化方式を中心に、圧縮されたビットストリームの生成及び転送の仕組みについて解説する。

**キーワード** MPEG, 動画像符号化, MPEG-4 AVC, H.264

## [Tutorial Invited Lecture] MPEG Video Coding System

Hiroshi WATANABE<sup>†</sup>

<sup>†</sup> Graduate School of Global Information and Telecommunication Studies, Waseda University

1011 Nishitomida, Honjo, Saitama 367-0035 Japan

E-mail: <sup>†</sup>hiroshi.watanabe@waseda.jp

**Abstract** MPEG is an abbreviated expression of “Moving Picture Experts Group.” It means both standard developing group and specification itself for audio-visual representation at ISO/IEC JTC 1. Major MPEG standards are MPEG-1, MPEG-2, MPEG-4 and MPEG-4 AVC. Among them, MPEG-2 is used for DVD and HDTV and MPEG-4 AVC is used for one-segment mobile TV broadcasting, Blu-ray Disk video coding and YouTube. In this report, mainly MPEG-2 video coding system is introduced since it certainly consumes a lot of network bandwidths and how compressed bitstreams are generated and transported is described.

**Keyword** MPEG, Video Coding, MPEG-4 AVC, H.264

### 1. まえがき

MPEG は Moving Picture Experts Group の略称であり、ISO/IEC JTC1 におけるオーディオビジュアル符号化標準の開発機関及び規格の両者を意味する。MPEG ビデオ規格の代表的なものに MPEG-1, MPEG-2, MPEG-4, MPEG-4 AVC がある。

このうち、MPEG-1 は Video CD など蓄積メディアをターゲットとして開発された。通信用の符号化標準の開発は ITU-T の担務であるため、ISO/IEC では蓄積メディア用としての必要条件を満足する符号化標準の開発を目指した。当初、MPEG-1 は PC 上で動作するデジタル映像として流通した。ファイルの拡張子には \*.mpg が用いられた。しかし現在では MPEG-4 AVC な

どの後発の符号化方式の方が効率が良いためあまり使われていない。

MPEG-2 は MPEG-1 の仕様に加えて、高レート化及びインタレース画像対応を特徴とする汎用的な映像符号化方式である。MPEG-2 の主要ターゲットは DVD , HDTV である。一方、MPEG-4 AVC は MPEG-2 をさらに効率改善した方式であり、携帯端末用ワンセグ放送や BD 用映像符号化、YouTube などでも用いられている。

本稿では、ネットワーク帯域の消費量が大きいと考えられる MPEG-2 のビデオ符号化方式を中心に、圧縮されたビットストリームの生成及び転送の仕組みについて解説する。

## 2. MPEG-2 符号化方式

### 2.1. ヘッダ

MPEG では単一の画像フォーマットを規定しない。符号化データのヘッダに続く部分に水平・垂直・時間解像度をフラグとして記入する柔軟な方法を採用している。このフラグ情報を解読することにより、特定の復号後の画像が特定のハードウェアで表示できるかどうかの判別が行える。代表的な画像サイズとしてアスペクト比 16:9 の HDTV に対しては 1920 画素×1080 ライン、4:3 の従来のテレビ画像に対しては 2 タイプがある。図 1 に示す NTSC に基づいた 720 画素×480 ライン×29.97 frame/sec と、PAL に基づいた 720 画素×576 ライン×25 frame/s の 2 種類である。異なった画素アスペクト比に対処するため、画素の水平/垂直比のパターンをフラグ化してヘッダに多重化する。また映画のフレームレートも指定できる構成になっている。

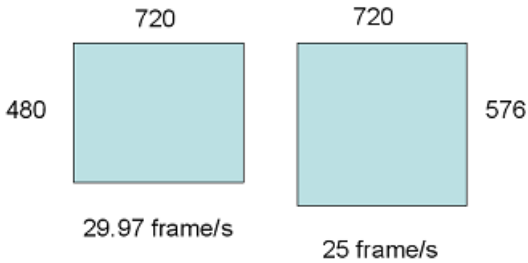


図 1 標準 TV に対する画像フォーマットの例

### 2.2. GOP

MPEG では GOP (Group of Pictures) と呼ばれる単位で符号化が行われる。GOP には、ランダムアクセス再生を可能とするため、動き補償フレーム間予測を全く用いずに符号化するフレーム (I ピクチャ) が挿入される。すなわち I ピクチャは静止画像として符号化される。I ピクチャの符号化方式は、ほぼ JPEG に近い手法が適用される。また、GOP では、蓄積/再生処理に要する時間的制約が少ないことを利用し、双方向動き補償によるフレーム間予測符号化を行うフレーム (B ピクチャ) を用いて、符号化効率を向上させている。通常の、順方向動き補償フレーム間予測符号化されるフレーム (P ピクチャ) は、B ピクチャの予測に用いられる。GOP 構造を図 2 に示す。

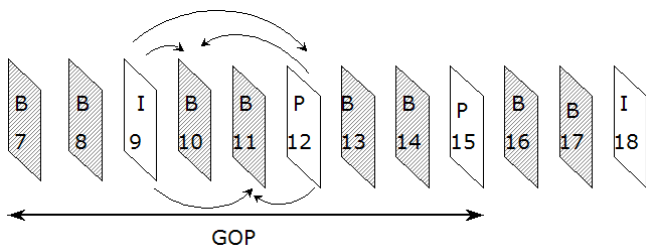


図 2 GOP (Group of Pictures) 構造

図 2 において、各フレームの番号は、入力時間順を示す。B ピクチャの符号化は P ピクチャの符号化後になるため、処理順序は、I→P→B→B→P→B→B→P→…のようになる。この構造により処理遅延が増大するが、隠れていた部分が現れる場合などに符号化効率を向上させることができる利点がある。

### 2.3. マクロブロック

MPEG では、動き補償フレーム間予測符号化が用いられる。動き補償を行うためには、動きベクトル検出をブロック単位に行う。このブロックサイズは 16 画素×16 ラインであり、マクロブロックと呼ばれる。マクロブロックは、輝度ブロックだけでなく 2 つの色差ブロックも含む。色差信号は、輝度信号に比べて、水平および垂直方向に半分の解像度を持ち、ブロックサイズは 8 画素×8 ラインとなる。得られた動きベクトルは、右および下方向を正と定義する (X, Y) の 2 次元ベクトルとして扱う。動き補償フレーム間予測による差分データは 8×8 画素単位で DCT により変換されて、符号化される。一方、動きベクトルは、マクロブロック単位で見ると隣接マクロブロックのベクトルと近い値をとることが多くなる。そこで、動きベクトルの符号化には、隣接マクロブロックの動きベクトルの値で予測した後、予測差分ベクトルをハフマン符号化する。ここで、動きベクトルの差分は、水平および垂直方向のベクトルについてそれぞれ独立に計算する。一般に、画像内で多くの領域が同じ方向に動いている場合、予測差分ベクトルは 0 近傍に集中する。そこで、0 近傍の差分ベクトルに対する符号を短く設定することにより、符号化効率の向上を図っている。

### 2.4. シンタクス

符号化データであるビットストリームの構文をシンタクスと呼ぶ。シンタクスは、階層構造になっており、シーケンス、GOP、ピクチャ、スライス、マクロブロック、ブロックの 6 階層で多重化される。シンタクスのレイヤと機能を表 1 に示す。テレビ会議用のビデオ符号化標準である H.261 や H.263 のビットストリームは通信を基本としているため、ビット単位でスタートコードの位置が変化する。しかし、MPEG では、ディスクからのピックアップ時の読み取り単位を考慮してスタートコードはバイト単位に揃えられている。

シーケンスは、そのヘッダを除き、GOP の繰り返りで構成される。GOP は I, P, B ピクチャの組み合わせからなり、ピクチャは垂直方向の 16 ラインからなる複数のスライスで構成される。スライスは 1 つ以上のマクロブロックからなり、マクロブロックは、4 個の輝度信号ブロックと 2 個の色差信号ブロックからなる。4:2:2 フォーマットや 4:4:4 フォーマットの場合には色差ブロックの数が異なる。

表 1 シンタクスのレイヤと機能

シンタクスのレイヤ	機能
シーケンス	プログラム内容のランダムアクセス単位
GOP	ビデオのランダムアクセス単位
ピクチャ	基本的な符号化単位
スライス	同期回復単位
マクロブロック	予測単位
ブロック	DCT 単位

図 3 に、MPEG ビデオシーケンスの空間的な階層構造を示す。これらの階層構造は、同期を取るためのユニークなヘッダ（可変長符号の中にあつて、特殊で見つけやすい符号）の確保が第一の目的であるが、階層化による処理の並列化も可能にしている。すなわちあるレイヤ以下は同じような処理が繰り返されるため、ソフトウェアやハードウェアを専用設計できる利点がある。

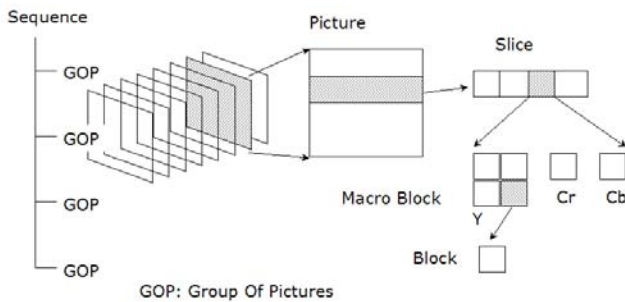


図 3 MPEG ビデオシーケンスの空間的な階層構造

### 2.5. インタレース対応

プログレッシブ画像を対象とする MPEG-1 とは異なり、MPEG-2 では、現在の標準 TV の信号形式であるインタレース画像を扱えることが特徴である。インタレース画像は NTSC 信号の場合、29.97Hz のフレームからなり、1 フレームは 59.94Hz の 2 フィールドからなる。これらのフィールドは、垂直方向のライン位置が交互になっている。インタレース画像は被写体が動いている場合には、垂直方向の解像度を半分にして、動きの再現性を高めるように動作する。逆に動きがない場合には、垂直方向の解像度が 2 倍になる。

MPEG-2 では、インタレース画像に対して、効率良く動き補償フレーム間予測を行うための工夫がなされている。動き補償フレーム間予測を行うための、基準画像の取り方として、フレーム構造とフィールド構造がある。フレーム構造では、2 つのフィールドを用いてフレーム単位あるいはフィールド単位に動き補償を行う。フィールド構造では、動き補償による予測の対

象は常にどちらかのフィールドであり、フィールド単位の動き補償に限定される。

フレーム構造の場合には、フレーム単位の予測を用いることができる。順方向予測と逆方向予測を組み合わせる双方向予測を、前後のフレームから行うことができる。それに加えて、対象とする K フレームのマクロブロックの信号を 2 つのフィールドに分割して、動き補償を行うフィールド単位の予測も可能である。フィールド単位の処理では、1 つのフレームは、トップフィールドとボトムフィールドに分割される。それぞれのフィールドは、前後のフレームに含まれる 4 つのフィールドから動き補償を行うことができる。

フィールド構造の場合には、予測の対象は個々のフィールドであるので、フィールド単位の予測だけしか用いることができない。フィールド構造はフレーム構造より単純であり、ハードウェアの簡単化を目的としている。

フレーム構造とフィールド構造における動き補償フレーム間予測には、いくつかのバリエーションがある。フレーム構造の場合には、16×16 画素のフレーム単位の予測、16×8 画素のフィールド 2 個を組み合わせるフィールド単位の予測、さらに、デュアルプライムと呼ばれる予測を切り替えて用いることができる。

入力画像や動き補償フレーム間差分画像は、8×8 画素のブロック毎に DCT が適用され、係数に変換された後にハフマン符号化される。各ブロックでは、画像に含まれる被写体に動きがあるかどうかにより、垂直方向の画素の類似性が異なる。フィールド間に動きがない場合には、フィールド単位のブロックを形成するよりも、フレーム単位のブロックで DCT を適用した方が効率的である。隣接ライン間の画素の類似性を利用でき、相関が高いため、DCT 係数がより低域周波数に偏る。逆に、フィールド間に動きがあると、フレーム単位のブロックでは、隣接ライン間の画素相関が低下する。このような場合にはフィールド単位の DCT を用いた方が、圧縮率の点で有利となる

## 2.6. プロファイルとレベル

MPEG-2 は仕様用途を限定しない汎用符号化方式として、種々のユーザ要求を満足するように設計されている。例えば、低遅延、高画質、ランダムアクセス、スケーラビリティなどがある。要件に対応できるように、エンコーダの機能やパラメータ（例えば、画像サイズや量子化特性、符号化ビットレートなど）を選択可能としている。これらのパラメータを符号化データのヘッダ部分に記載する。デコーダでは復号に先立ち、ヘッダ部分を読み込んでパラメータを知り、復号動作を決定する。しかし、パラメータの選択の自由度が大きく、デコーダ設計が複雑になる。そこで、デコーダがどの範囲まで対応できるかを示す、機能・パラメータ群が定義されている。これを「プロファイル」と「レベル」と呼ぶ。プロファイルは、符号化ツールを適当な機能単位で区切ったものである。レベルは、対応できる画像サイズで区切ったものである。

メインプロファイル・メインレベル(MP@ML)は最も多くのアプリケーションで使われる互換ポイントである。DVD と通常のデジタルテレビ放送は MP@ML を用いる。また HDTV ではメインプロファイル・ハイレベル (MP@HL) を用いる。予測には、双方向予測を用いることができる。双方向予測には遅延が大きくなる欠点があるが、DVD などの遅延が問題とならない蓄積メディアのアプリケーションでは、符号化効率を優先させてこの予測方式が用いられる。メインプロファイルの仕様を表 2 に示す。

## 2.7. VBV (Virtual Buffer Verifier)

MPEG-2 ビデオにより生成されるフレーム毎のデータ量の例を図 4 に示す。I ピクチャでは予測符号化を用いることができないため、データ量は必然的に大きくなる。P ピクチャでは I ピクチャからの予測を用いることができるため、同程度の画品質を得るためにはデータ量は 1/4 程度で済む。B ピクチャでは双方向の予測が効果を発揮するため、さらに少ないデータ量となる。画像中のオブジェクトが平行移動だけの場合には、ほぼ動きベクトルだけのデータで画像を構成でき、DCT 係数を必要としないことになる。このような場合には B ピクチャのデータ量は極端に減少する。

フレーム毎に発生情報量が大きく変化するため、一定レートのチャンネルにデータを転送するためにはバッファメモリが必須となる。

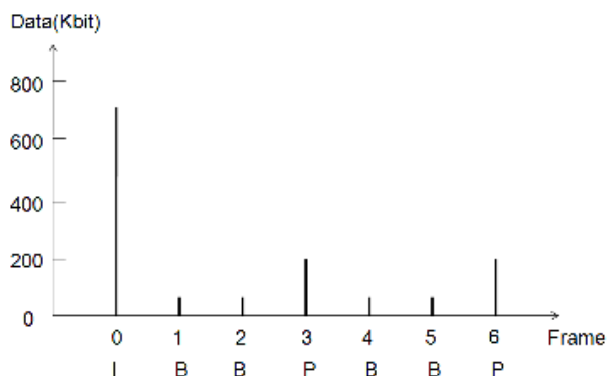


図 4 フレーム毎のデータ量の例

表 2 メインレベル・メインプロファイルの仕様

項目	内容
画像フォーマット	ITU-R601 サイズ以下 (720x480x29.97Hz, 720x576x25Hz)
符号化ビットレート	15Mbit/s 以下
色差形式	4:2:0
ピクチャタイプ	I, P, B
符号化構造	フレーム構造およびフィールド構造
動き補償フレーム間予測 フレーム構造の場合 フィールド構造の場合	フレーム 16x16, フィールド 16x8, デュアルプライム予測 フィールド 16x16, フィールド 16x8, デュアルプライム予測
動きベクトルの範囲	-127.5~127.5 画素, 0.5 画素精度
バッファサイズ	1 835 008 ビット以下
互換性	MPEG-1 互換 (MPEG-2 デコーダが MPEG-1 符号化ビット ストリームを復号可能)
イントラ DC 係数予測	10 ビット以下
イントラ VLC	MPEG-1 および新テーブルをピクチャレイヤで選択
DCT 係数スキャン	ジグザグおよび新スキャンをピクチャレイヤで選択
エラー耐性	イントラマクロブロックに動きベクトル付加
VBR モード	含まれる

復号器（デコーダ）側では、一定転送速度のチャンネルからデータを受けてバッファメモリに蓄えた後、復号を開始する。その際にバッファメモリが空になったり、オーバーフローしたりしないことが重要である。そこで符号化器（エンコーダ）では、仮想的に復号バッファ検証器(Virtual Buffer Verifier)を持ち、1フレーム毎の符号化ビット数がオーバーフローやアンダーフローを起こさないように符号化制御を行う。どれだけバッファにデータが溜まった時点で復号を開始すれば良いかを示す値はヘッダに書き込まれる必要がある。この値が大きいとストリーミングなどの Web アプリケーションでは、データを受信してから再生までの時間が長くなるという問題を生じる。

### 3. MPEG-4 AVC (H.264)

#### 3.1. 共同規格

MPEG-4 AVC, は ISO/IEC JTC 1 と ITU-T の共同作業グループ(Joint Video Team: JVT)により 2003 年に作成された動画符号化標準(ISO/IEC 14496-10)である。特徴は圧縮率を高めたことにあり、ISO/IEC における名称は高度ビデオ符号化(Advanced Video Coding)であり、ITU-T では、H.264 と呼ばれる。AVC は MPEG-4 の Part10 という位置付けであるが、いわゆる MPEG-4 ビデオ(MPEG-4 Part2)の特徴であるオブジェクトベースの符号化ではなく、符号化効率の向上を目指したものである。また、H.264 は ITU-T 側から見れば、テレビ電話用に開発された H.261, H.263 を継承する高性能なビデオ符号化である。

#### 3.2. 符号化効率

AVC は通常の MPEG 符号化に比べて 1.5–2 倍の符号化効率を達成できる。基本的な符号化アルゴリズムは、従来の MPEG と同様にマクロブロック単位の予測や直交変換や量子化を用いる。しかし、予測や変換に使われる符号化ツールは非常に数多く、それらの中から最適なものを選び出して使う仕組みをとっている。一方、新しいエントロピー符号化やブロック除去フィルタは従来の MPEG 符号化にはなかった技術要素であり、符号化効率すなわち同ビットレートでの画質改善に大きく貢献している。特に MPEG-2 に比べて、可変ブロックサイズ動き補償、重み付き予測、INTRA 予測、デブロッキングフィルタ、エントロピー符号化などが性能改善に大きく貢献している。

#### 3.3. スライス

AVC では 1 フレームの画像の中で異なった種類の予測をスライス単位で用いることができる。スライスには、I-スライス、P-スライス、B-スライスがある。I-スライスでは、 $16 \times 16$  画素のブロック単位に周辺の画素値を用いて 4 種類の外挿予測を行う。また  $16 \times 16$

画素のブロックを 4 分割し、 $4 \times 4$  画素のブロック単位に 9 種類の外挿予測を切り替えて用いることができる。これらは INTRA 予測と呼ばれる。

P-スライスでは、可変ブロック動き補償フレーム間予測を行う。 $16 \times 16$  画素のブロックを、 $16 \times 8$ ,  $8 \times 16$ ,  $8 \times 8$  画素の 4 種類に分割して処理できる。 $8 \times 8$  画素のブロックは、さらに、 $8 \times 4$ ,  $4 \times 8$ ,  $4 \times 4$  画素の 4 種類のサイズに分割できる。

B-スライスは、必ずしも時間的に過去と未来のフレームから予測するのではなく、過去の 2 つのフレームから加重平均して予測画像を生成することができる。このため、フェードインやフェードアウトといった徐々に輝度に変化するような場合にも、効率良く予測が行える。

動きベクトルは  $1/4$  画素精度で計算される。 $1/2$  画素精度のための画素値の計算には 6 タップの FIR フィルタが用いられる。次に  $1/4$  画素精度のための画素値の計算には、縦横あるいは斜め方向に 2 タップの FIR フィルタが用いられる。P-スライスの予測に用いる画像は、既に符号化した画像のなかから、いくつかの画像を指定して行われる。このために、符号化済みの画像に対して、予測に用いるかどうかを示す参照ピクチャリストが準備される。

#### 3.4. 可変ブロックサイズ動き補償

AVC における変化の最小単位は、予測の最小単位である  $4 \times 4$  画素のブロックである。さらに、実数演算に起因する DCT と IDCT のミスマッチを避けるために、DCT を近似した変換を 2.5 倍して整数化した変換 T が用いられ、整数精度 DCT と呼ばれる。また、逆変換 T-1 も対応する整数変換で与えられる。

また、イントラモードの  $16 \times 16$  画素のブロックについては、 $4 \times 4$  の整数変換により 16 個の DC 係数が得られる。これらの DC 係数に対してはアダマール変換を用いる。色差成分 Cr, Cb のブロックについては、サイズが縦横  $1/2$  であるため、それぞれ 4 個の DC 成分に対してアダマール変換を用いる。

#### 3.5. エントロピー符号化

エントロピー符号化は AVC において特に大きく変わった点である。改良型のハフマン符号化の他に、算術符号化が取り入れられた。また、ピクチャヘッダなどの重要な情報には、誤り耐性の強い指数ゴロム (Golomb) 符号化が用いられる。指数ゴロム符号は、符号の先頭から、プリフィックス (数個の符号 “0”), セパレータ (符号 “1”), サフィックス (数個の “1” と “0” の組み合わせ) からなる。プリフィックスとサフィックスの長さは同じである。最も短い符号は “1” であり、ビット数を増やすに従い、順次 “010”, “011”

と続く。これらを非負整数値に対応させる。

指数ゴロム符号を使用しない部分に対しては、改良型ハフマン符号化である CAVLC (Context-based Adaptive Variable Length Coding, コンテキスト適応型可変長符号化) と、算術符号化の CABAC (Context-based Adaptive Binary Arithmetic Coding, コンテキスト適応型 2 値算術符号化) が用いられる。AVC のメインプロファイルでは、これらはピクチャ単位に切替え可能であるが、ベースラインプロファイルや拡張プロファイルでは、CAVLC のみ使用できる。

### 3.6. AVC プロファイル

AVC には表 3 に示す基本的な 3 つのプロファイルがあり、予測に用いるスライスの種類、エントロピー符号化、インタレース対応の観点から決められる。レベルは 15 種類が符号化ビットレート (64kbit/sec - 240Mbit/sec) に応じて設定される。

ベースラインプロファイルは、エラー耐性に優れており、テレビ電話やテレビ会議、携帯電話などに用いられる。メインプロファイルは、高圧縮用であり、放送や蓄積メディアに適する。拡張プロファイルはインターネットでの利用を目的とする。特にベースラインプロファイルは、既にテレビ会議装置や地上波デジタル放送 (ワンセグ) に採用されている。

表 3 AVC におけるプロファイル

プロファイル	スライスの種類	エントロピー符号化	インタレース対応
ベースライン	I, P-スライス	CAVLC	不可
拡張	I, P, B-スライス	CAVLC	不可
メイン	I, P, B-スライス	CAVLC, CABAC	可

## 4. アプリケーション

### 4.1. TS と PS

テレビジョン放送や通信で用いられる MPEG-2 は蓄積メディアとは異なった多重化が行われる。オーディオデータや他のメタデータと共に多重化されたビデオデータは、伝送目的の場合トランスポートストリーム (Transport Stream; TS) という形式に整形される。一方、蓄積メディアではプログラムストリームと (Program Stream; PS) という形式に整形される。トランスポートストリームは 188 バイトの小さなパケットであり、オーディオデータが一定間隔で挿入される。データを受信次第、順次復号再生するアプリケーションに適している。プログラムストリームは 1 フレーム分の圧縮画像データとオーディオデータを並べて、大きなパケットとして扱う。遅延が許容できる蓄積メディアのアプリケーションに適している。

### 4.2. エラー耐性

通信時のネットワークエラーや DVD からの転送エラーが生じた場合、ビットエラーに近ければ映像は基本的にスライス単位で壊れる。復号側の機能として前フレームの画像で補完したり、動きベクトルを用いてエラーを目立たなくしたりする工夫が可能である。しかしバーストエラーのような場合には、次の I ピクチャ或いは I スライスまで正常な復号が不可能となる。そのため放送などでは GOP サイズを 15、すなわち 1/2 秒毎に回復可能に設定している。

### 4.3. YouTube

YouTube では低ビットレート向けの独自の符号化方式 On2 に代えて、MPEG-4 AVC が使用され始めている。ビデオをアップロードする際に HDTV レベルの高精細、高品質なデータを送っても、通常のスリーミング時には低品質な映像として提供される。しかし、サーバ内部には高品質なバージョンも存在している。URL の最後尾に &fmt=22 などアップロード時のフォーマットに対応した英数字を追加すると、元の品質での画像再生が可能となる。このオプションを多くのユーザが利用するとネットワークのトラフィックが急増する可能性がある。

## 5. むすび

本稿では、MPEG ビデオ符号化方式について、ネットワーク帯域の消費量が大きいと考えられる MPEG-2 のビデオ符号化方式を中心に、圧縮されたビットストリームの生成及び転送の仕組みについて解説した。

### 文 献

- [1] 亀山 渉、花村 剛、“改訂版 デジタル放送教科書 (上) MPEG-1/2/4~H.264/AVC”, インプレス標準教科書シリーズ, Sep. 2004.