

新ビジネスを拓く H.264/AVC

その技術の中核をとらえる

2003年に標準化が終了した動画圧縮技術「H.264/AVC」は、その優れた性能により幅広い分野で利用されつつある。すでに、本誌の特集記事で概要を紹介したが、ここではこの技術を、MPEG-1/2などの既存の規格とも比較しつつ、技術的側面から解説する。なお、記事末に用語解説を設けたので活用してほしい。

鈴木 輝彦

ソニー シニアシグナルプロセッシングリサーチャー

ハイビジョン時代の 新圧縮技術「H.264/AVC」

1990年にISDN回線に対応した「H.261」というマルチメディア通信用の圧縮規格が、テレビ会議用の動画圧縮方式として最初に標準化されて以来、動画圧縮方式はその圧縮効率を改善し、図1のように進化し続けてきた。さらに、1995年に汎用的な目的を持ったMPEG-2が標準化され、DVDのような蓄積メディアや、放送、通信などの幅広い分野で使用されるようになり、急激に普及してきた。その後、H.263、MPEG-4などの圧縮方式が

制定され、さらに2003年には、注目される次世代圧縮方式「H.264/AVC」(Advanced Video Coding、高度画像符号化)の標準化が終了した。

MPEG-2の標準化終了以来、8年を経て規格化されたこの汎用的な新圧縮方式は、MPEG-2と比較すると、ほぼ2倍の圧縮効率を達成することができた。

時代は、3Gモバイル時代、デジタル放送時代、ハイビジョン(高精細テレビ)時代を迎えており、小型モバイル画面で高画質を、パソコン画面で高画質を、大型テレビ画面で高画質を実現できるという「HDTV(ハイビジョン)時代」の幅広い

ニーズに応えた圧縮技術となっている。

2003年3月にH.264/AVC第1版(バージョン1)が発行された後、より高画質で業務用まで使用できるFRExt(Fidelity Range Extension、原画をより忠実に再現できるようにするための機能の拡張)の標準化が進められ、2004年4月にその標準化作業を終了した。

放送からゲームまで、 予想以上のインパクト

標準化作業終了後、産業へのインパクトは予想以上に強く、図2に示すように、さまざまなコンピュータ・放送・通信などに関連する団体や組織がH.264/AVCの正式な採用を決定している。

当初アプリケーションとして想定されていたテレビ会議システムでは、すでに主要メーカーはH.264/AVCに対応済みである。また、日本における1セグメント放送をはじめ韓国のDMBや欧州におけるデジタル放送であるDVB、次世代蓄積メディアであるBlu-ray DiscやHDDVD、第3世代携帯電話通信方式の標準化団体である3GPPもすでに正式採用を発表、さらにアップルコンピュータもQuickTimeへの標準搭載を発表している。2004年12月に発売された次世代小型ゲーム機「PSP」(Play Station Portable)にも採用され、その画質の高さが注目されている。

今後は、図2に示すような幅広い産業の画像アプリケーションでH.264/AVCが主流になり、新しいビジネスが期待され、すでに胎動し始めている。

次に、H.264/AVCがどのようにして、

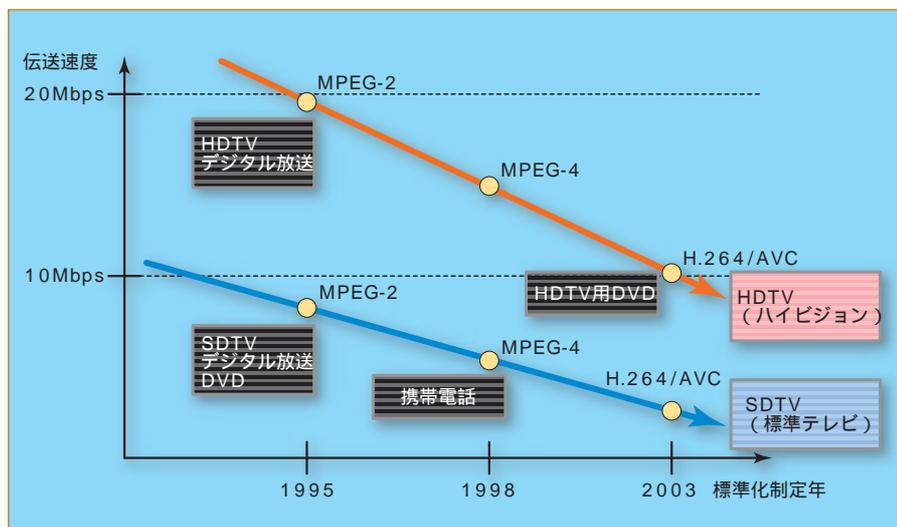


図1 動画圧縮方式の歴史と圧縮効率の改善

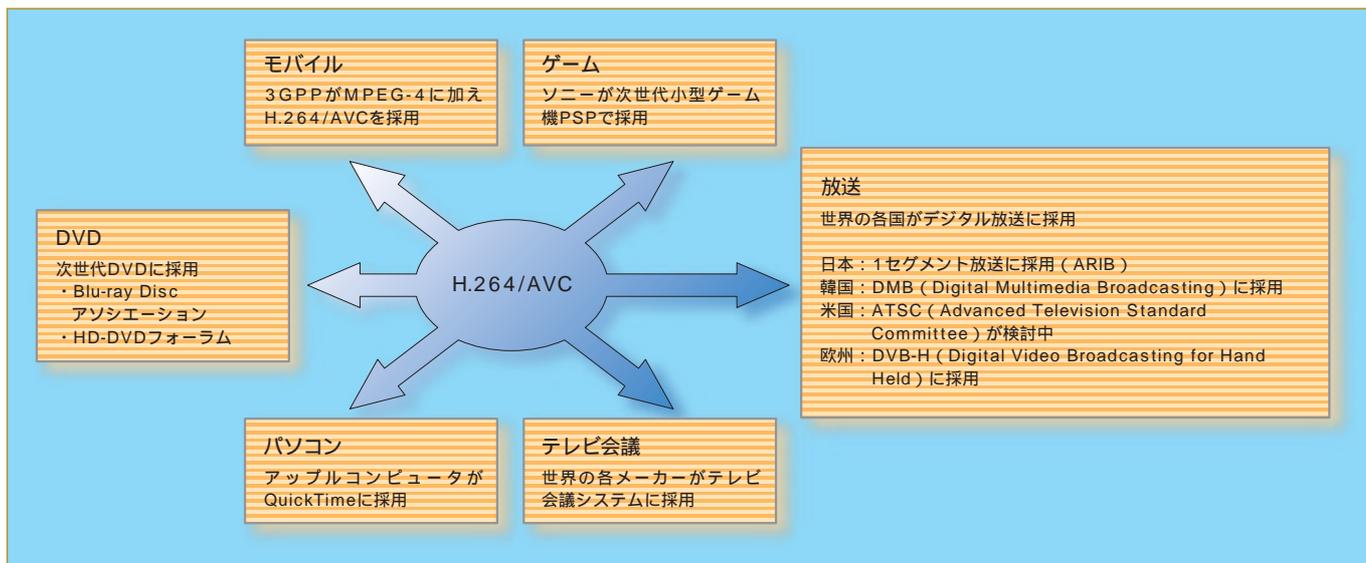


図2 幅広く汎用的に採用されている H.264/AVC

高い圧縮率を実現できたか、その秘密(中核技術)を簡単に紹介しよう。

H.264/AVCのコア技術

〔1〕圧縮効率改善の要因：

新技術と既存技術の最適化

前述したように、H.264/AVCは既存の圧縮規格(MPEG-2など)の2倍の圧縮効率を目標に標準化が進められてきた。圧縮のために画像を符号化した(Encoder)圧縮された画像を元に戻す「復号」した(Decoder)するコーデック(CODEC = Coder + Decoder、符号化/復号装置)の構成自体は、従来の圧縮技術である「動き補償 + DCT(コラム参照)」を用いた方法が引き続き採用されている。

すなわち、H.264/AVCは、新規の技術だけで2倍の圧縮効率を達成しているのではなく、既存のMPEG-2/4などで使用されている、圧縮のための個々の要素技術(ツール)をより最適化し、それらをそれぞれ足し合わせ(積み重ね)ることによって、2倍の圧縮効率を達成している。

〔2〕圧縮効率の改善を実現した

主なコア技術

このように、H.264/AVCが高い圧縮を実現できたのは、圧縮効率を改善できた

ことであり、その主な要因は、予測効率の改善にある。ここで、「予測」とは、今まで(過去)の流れから次にどのような画面がくるかを予測し、過去の画面との差分だけを送って(似ている重複部分は送らない)結果として画像を圧縮するという意味である。この予測を効率よく行うことが、予測効率の改善となる。

やや専門的になるが、H.264/AVCで

は次に説明する(1)~(6)などによって高い圧縮率を実現している。

(1)イントラ符号化における

予測効率の改善

画像圧縮の方法の主要な1つである、1枚の画面(フレーム)内の画像を圧縮する方法は、フレーム(画面)内符号化あるいはイントラ符号化と呼ばれている。例

圧縮技術のポイント：「DCT」「DCT係数」とは？

DCTとは、Discrete Cosine Transformの略で、離散コサイン変換と訳される(離散とは「飛び飛びの」、すなわち「不連続」という意味)。

DCTは圧縮技術ではなく、元の画像を圧縮しやすくするために重要な、数学的変換の仕組みのことである。例えば、元の画像を圧縮しやすくするために、元の画像を構成する「画素値」を別の情報である「周波数成分」に変換することによって、画像の冗長部分を削減しやすくする方法である(冗長部分とは、削除しても元の画質を損なわない程度の情報のこと)。

画像にDCTを適用して変換(画素値を周波数成分に変換)して得られた値は、「DCT係数」といわれる。

DCT係数は、

周波数がゼロのDC成分(直流成分。DC：Direct Current、画像がまったく変化しない部分)と、

低周波(画像の変化が小さい部分)から高周波(画像の変化が大きい部分)を含むAC成分(交流成分。AC：Alternate Current)の2つで構成される。

一般に自然画では、画像の変化は小さくなるからであるため、DCT係数はほとんどDC成分と、AC成分の内の低周波成分に偏って分布する。

このため、高周波成分は少ない(削除しても画質に影響を与えない)ので、例えば高周波成分を削除して、画像を圧縮することができる。

例えば、標準テレビ(SDTV)の1枚の画面は720画素×480ライン=34万5600画素の解像度を持っているが、イントラ符号化とは、この画面の中の画素から冗長な画素をどれだけ削除して符号化できる(圧縮できる)か、ということである。

既存のH.261のイントラ符号化の場合、予測をまったく使用しないで符号化を行っている。これはリアルタイムなテレビ会議を行う際の伝送時のエラーを考慮して、できるだけ予測を行わないようにしているためである。

MPEG-2では、DCT変換を行った結果得られるDCT係数のDC成分(画像が変化しない部分)のみ、周囲の画像ブロックから予測を行い、その差分を符号化している。MPEG-4およびH.263では、さらに進めてDCTのDC成分に加えて、低域(低周波数)のAC成分(画像の変化した部分)を周囲の画像ブロックから予測している。このようにMPEG-2、MPEG-4、H.263では、DCT変換を行った結果として得られるDCT係数を、周囲の画像ブロックから予測するアプローチをとっていた(DCT係数は周囲の画素ブロックとの間で相関があまりないため、予測効率が悪いことが問題であった)。

これに対してH.264/AVCでは、DCT変換してから予測するのではなく、相関

のある周囲の画素から予測してから差分信号をDCT変換して、圧縮効率を高めている。

(2)動き補償における予測効率の改善

動き補償とは、直前の画面内における被写体の動きの量を検出して、効果的な予測画面をつくる手法のことである。この動き補償を行う際の処理単位を「動き補償ブロック」という。

H.264/AVCにおいては、動き補償を行う際、画像の処理単位である1つのマクロブロック(16×16画素)を、図3に示すように16×16画素、16×8画素、8×16画素、8×8画素の動き補償ブロックに分割し、この単位で動き補償を行うことが可能である。さらに、8×8動き補償ブロックに関しては、8×8画素、8×4画素、4×8画素、4×4画素の動き補償サブブロックに細分化し、この単位で動き補償を行うことも可能である。

それぞれの動き補償ブロックは、独立した動きベクトルを持つ。動きベクトルとは、動き補償ブロックが移動した方向とその移動量を示す値である。動きベクトルの精度は1/4画素精度である。

(3)整数DCT変換による圧縮効率の改善

MPEG-2、MPEG-4、H.263などの既存

の規格では、実数精度の8×8ブロックサイズのDCT変換が採用されてきた。しかし、H.264では4×4ブロックサイズの整数変換が採用された。これは圧縮効率の改善というよりも、複雑度(Complexity)の削減が主な目的であるが、これによりDCTの精度を規定する必要がなく、また各社デコーダー間のミスマッチ(接続不能)も生じないという利点がある。

画面を16×16画素(マクロブロック)の単位で処理する「イントラ16×16モード(マクロブロックモード)で符号化されたブロックは、最初に4×4整数変換を行った後、図4に示すように、各ブロックからDC成分(直流成分)のみを取り出し、4×4DCブロックを構成する。この4×4ブロックに、さらに、DCTよりも簡易な数学的な変換(直交変換)方式であるアダマール変換(図4①)を施し、より効率を高めている。

色差信号(画像の色を表す信号)も同様に、4×4整数変換を行った後2×2DCブロックを構成し、これに2×2ブロックサイズのアダマール変換(図4②、③)を行う。

アダマール変換された2次元(X-Y軸)のDCT係数は、ジグザグスキャン(ジグザグ走査)によって直列の1次元の信号に変換され、エントロピー符号化される。

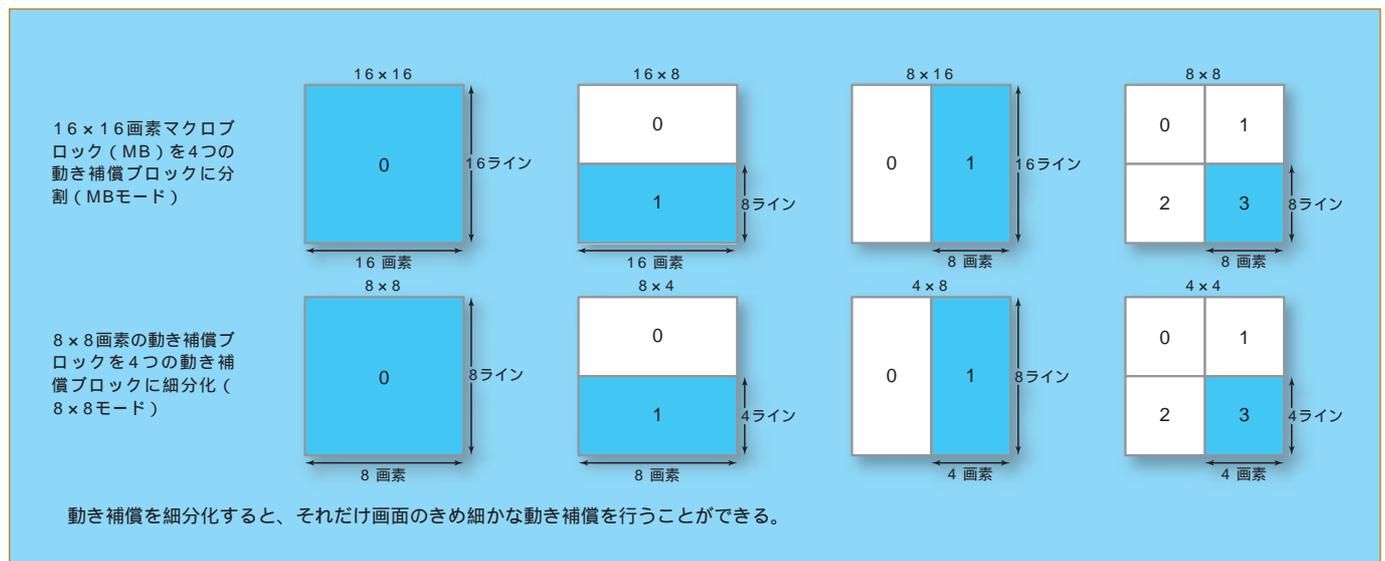


図3 H.264/AVCの動き補償ブロック

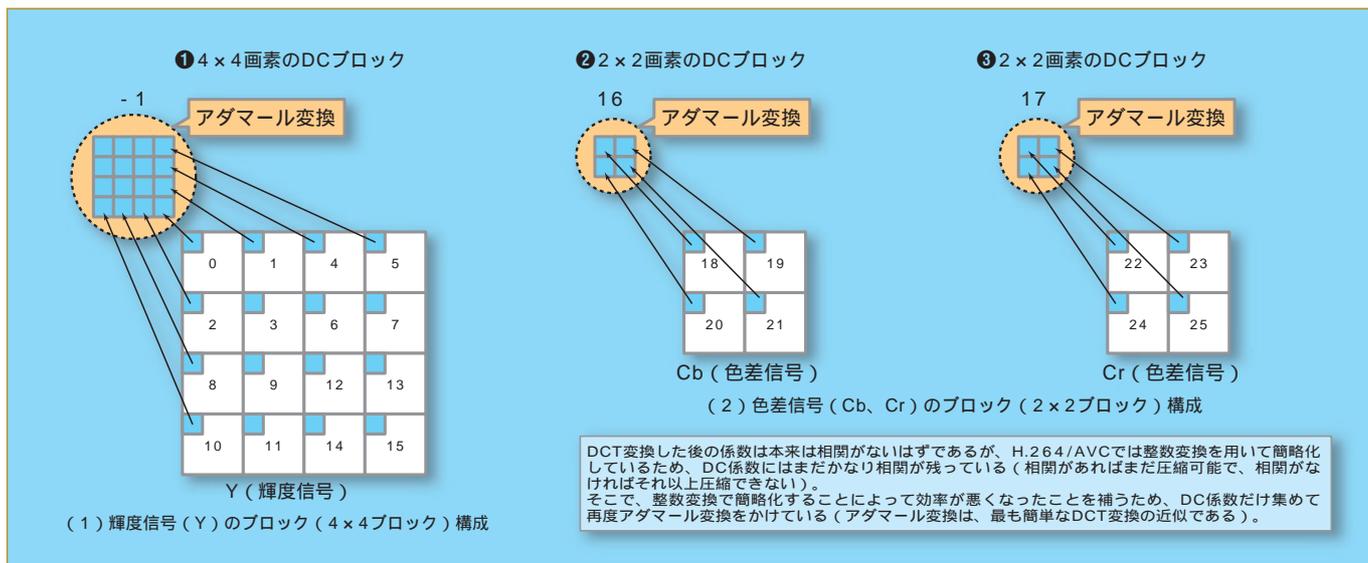


図4 イントラ16×16モードにおけるDCブロック(直流ブロック)の構成

エントロピー符号化とは、モルス符号のように、出現頻度の高い値には「短い符号」を、出現頻度の低い値には「長い符号」を割り振って圧縮効率を高める方式で、可変長符号化ともいわれている。

(4) エントロピー符号化における 予測効率の改善

H.264/AVCでは、各シンタクス要素を可変長符号(VLC: Variable Length Coding)またはCABACと呼ばれる算術符号化の2種類の方式によって符号化している。シンタクス要素の「シンタクス」とは、符号化されたデータ列の規則(構成方法)のことである。また、シンタクス要素とは、例えばシンタクスで伝送されることが規定されている、DCT係数のような変換係数や動きベクトルなどの情報のことをいう。

変換係数を可変長符号にする際には、CAVLCという可変長符号化が行われる。

一方CABACでは、各シンタクス要素ごとに、周囲の画素ブロックの状態に基づいて係数の発生状況(コンテキスト)を計算する。このコンテキストに応じて、確率モデルが定義されている。それぞれのシンタクス要素に最適な確率モデルに

基づいて符号化することによって、圧縮効率を改善している。

(5) デブロッキングフィルタによる 圧縮効率の改善

動き補償+DCT変換を採用する圧縮方式の場合、画素ブロックの境界で、人間の肩の部分などにギザギザができるなどの歪みが生じる。この劣化(歪み)を低減するために、既存方式の場合は、ポストフィルタを用いる。すなわち、圧縮されている画面のデコード(復号)が終了(ポスト)した後、画面を構成する各ブロックの境界にフィルタをかけて、歪みを平滑化して画質を向上させている。

一方、H.264/AVCでは、ブロック歪みを低減することを目的に、動き補償のループ(回路)内に、ブロック境界での平滑化を行い、符号化時に生じたブロック歪みを減少させる「デブロッキングフィルタ」が新しく導入された。

デブロッキング(ブロック歪み除去)フィルタで処理された画像は、出力画像として用いられるだけでなく、そのフレーム以降の動き補償の参照画像(基準画像)として、フレームを蓄積しておくフレームバッファ(メモリー)に保存される。これによって、参照画像がブロック

歪みのない滑らかな画像となっているため、圧縮効率や主観画質(論理的ではなく、実際に目で見えた画質)を向上させている。

(6) 第1版以降に追加されたFRExt

H.264/AVCの第1版では、4:2:0、8ビットのビデオ信号にしか対応していない。しかし、その後、H.264/AVCの高い圧縮特性を放送局などの業務用のアプリケーションにも活用したいという強い要望があり、第1版に引き続いて業務向けの拡張の標準化が行われた。この拡張では4:2:2、4:4:4、10~14ビットのビデオ信号に対応できるようになった。また、色空間も、DVDレコーダーなどで使用される「Y Cb Cr」のみではなく、テレビ画面に使用されるRGBや、デジタルシネマなどを想定して規格化されたXYZなどの色空間にも対応している。

当初は、こうした業務用の画像信号に対応することのみを目的としていたが、一方、前述した4:2:0、8ビットによる画像信号についても、より高精細に圧縮したいという要望があったため、4:2:0、8ビットビデオの高精細化もその作業に含まれることになった。名称も当初の

業務用拡張(Professional Extension)から高忠実度拡張(FRExt : Fidelity Range Extention)へ変更された。この結果、4 : 2 : 0、8ビット用に、8 × 8ブロック整数変換、量子化マトリクスなどのツールが、新たに採用された。

8 × 8 整数変換を用いることによって、細かいテクスチャー(絵柄)や映画素材のフィルムグレインノイズ(粒状ノイズ。フィルムの粒子を再現する非常に細かいノイズ)をより忠実に表現する(映画によっては、わざと粒状ノイズを出して古い時代の雰囲気を出すことを重視する作品もある)ことが可能となった。

また、量子化マトリクスにより、量子化を主観に基づいて重みを付けて行く決まった係数ではなく、必要に応じて重み付けした係数を掛け算することが可能になった。こうしたツールは、特に、HDTV 以上の高精細ビデオにおいて映画素材を圧縮する際に有効である。

H.264/AVC の プロファイルとレベル

[1] H.264/AVC の第 1 版の場合

画像の圧縮技術では、

- (1) プロファイル : 圧縮符号化の機能。
ツールの組み合わせ
- (2) レベル : 画像の解像度

が重要であり、これに関する規定がされている。H.264/AVC の第 1 版では、図 5 に示すように、ベースライン、メイン、拡張の 3 つのプロファイルが定義されている。

- (1) ベースラインプロファイルは、B ピクチャを含まない低遅延が要求されるアプリケーション用のプロファイルであり、テレビ会議用などが想定されている。
- (2) メインプロファイルは、ベースラインプロファイルや算術符号化(CABAC)をサポートし、放送や高画質なエンターテインメント用のプロファイルである。
- (3) 拡張プロファイルは、エラー耐性のツールなどを含み、主にストリーミングを想定したプロファイルである。

これまでの MPEG の規格では、上位のプロファイルは下位のプロファイルを含むようにオニオンリング(玉ねぎの輪)状に定義されてきた。しかし、H.264/AVC では、3 つのプロファイルでサポートするツールが異なるため、完全なオニオンリング状のプロファイル構造とはなっていない。ベースラインプロファイルのツールからエラー耐性のツールを除いたものが、各プロファイル間で共通となる。ただし、現在のところ、エラー耐性のツールを使用するアプリケーション

はないため、実質的にはオニオンリング構造となっている。

H.264/AVC のレベルは、現在 QCIF (180 × 144 画素) CIF (360 × 288 画素) SDTV (720 × 480 画素) HDTV (1920 × 1080 画素) SHDTV (1080P : 1920 × 1080 画素 × 60 Progressive Frame/second) の 5 つのレベルが定義されている。

[2] FRExt で

4 つのプロファイルを追加

FRExt では、メインプロファイルを拡張して新たに 4 つのプロファイルを定義した。図 6 に、FRExt で新たに定義されたプロファイルを示す。

4 つのプロファイルは、メインプロファイルをもとにして、オニオンリング構造に定義されており、それぞれ下位のプロファイルのビットストリームを復号できなければならない。

- (1) High プロファイルは、メインプロファイルに 8 × 8 整数変換、量子化マトリクスを追加したもので 4 : 2 : 0、8 ビットのビデオ信号のみサポートする。HDTV 用の蓄積メディアや放送などに利用されることを想定している。次世代 HDTV 用光ディスクでは、この High プロファイルを採用することが決定された。

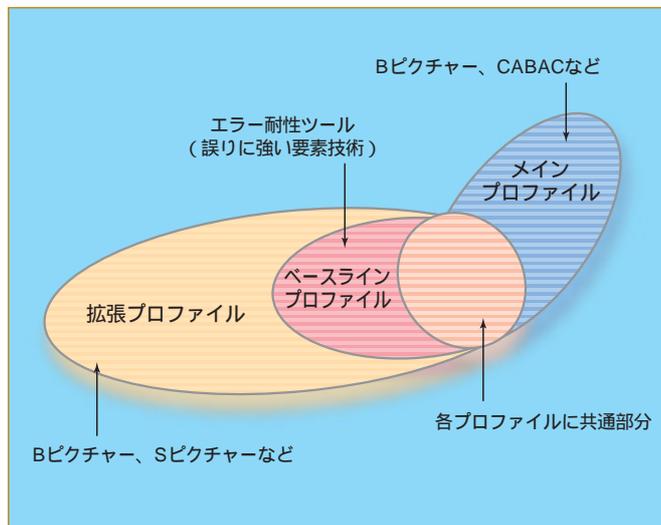


図 5 H.264/AVC 第 1 版(バージョン 1)におけるプロファイル

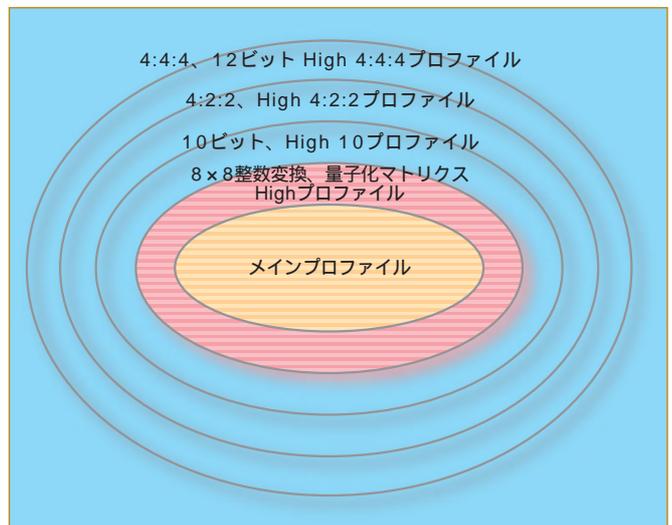


図 6 FRExt での新しい 4 つのプロファイル(メインプロファイルの拡張)

- (2) High 10 プロファイルは、High プロファイルに追加して 10 ビットビデオをサポートする。次世代のコンシューマーモニターは 10 ビットのビデオ信号をサポートすることから、将来のコンシューマー用途を想定したプロファイルである。
- (3) High 4 : 2 : 2 プロファイルは、4 : 2 : 2、10 ビットビデオをサポートするプロファイルである。ほとんどのスタジオで使用されている業務用機器は 4 : 2 : 2、10 ビットのビデオ信号を使用しているため、業務用の主力となることを想定したプロファイルである。
- (4) High 4 : 4 : 4 プロファイルは、4 : 4 : 4、12 ビットビデオまでサポートしたプロファイルで、デジタルシネマやハイエンドの業務用機器に使用されることを想定している。

H.264/AVC 用語解説

3GPP	Third Generation Partnership Project、第 3 世代パートナーシップ・プロジェクト。第 3 世代移動体通信方式の標準化組織の 1 つ。	量子化	アナログの連続した元の画像を、飛び飛びのデジタルの画像に変換する操作のこと。量子化するビット数が大きいほどきめ細かい画像が得られる。例えば、4 ビット(16 レベルの量子化。量子化レベルが低い)よりも 8 ビット(256 レベル。量子化レベルが高い)の方がきめ細かい画像となる。
符号化	符号化とは、例えばデジタル情報を DVD に蓄積する場合、DVD に蓄積するために都合のよい形(これを符号という)に変換すること。通常、その符号化の際に情報が圧縮されるので、「符号化する」ことを「圧縮する」と同義語に使う場合が多い。	量子化マトリクス	それぞれの DCT 係数に異なる重み付けを与える仕組みがあるが、量子化マトリクスは、これを行うために用意されたテーブル(一覧表)のこと。
CABAC	Context-Adaptive Binary Arithmetic Coding、コンテキスト適応型 2 値算術符号化方式。コンテキスト適応型とは、周囲の状況(コンテキスト)に応じて効率の良い符号化方式を適応的に選択する方式のこと。算術符号化方式とは、可変長符号化のように、符号化する対象が固定しているとして符号を決めるのではなく、時間とともに変化する符号化の対象の統計を計り、これに適応しながら符号化して、符号化効率を高める方法。	B ピクチャー	Bi-predictive Prediction-Picture、双予測ピクチャー。ピクチャーとは 1 枚の画面データのこと。画像圧縮では、I ピクチャー(Intra Picture)、P ピクチャー(Predictive)、B ピクチャーの 3 種類のピクチャーが規定されている。I ピクチャーは、その画面内だけで独立して符号化して得られる画面、P ピクチャーは前(過去)の画面(順方向)から予測して作られる画面、B ピクチャーは過去の 2 枚あるいは未来の 2 枚の画面から予測して作られる画面(注。MPEG-1/2/4 の B ピクチャーでは、各 1 枚からの予測画面であった)。
CAVLC	Context Adaptive Variable Length Coding。コンテキスト適応型可変長符号化方式。可変長符号化方式の一種で、周囲のブロックでの係数の発生状況(コンテキスト)に応じて可変長符号を切り替え、より適切な可変長符号を用いて符号化することによって符号化効率の改善を図る方式。	S ピクチャー	Switching Picture、ストリームの切り替え(スイッチング)を行うピクチャー。H.264/AVC で採用された新しいピクチャー。
4 : 2 : 0、8 ビット	民生用のビデオに使用されている、ビデオ信号のフォーマット。輝度信号に対して色情報は、水平、垂直方向ともに半分の解像度。各信号は 8 ビットである。これ以外のフォーマットはすべて業務用に用いられる。	マクロブロック	画像を圧縮する際の処理単位をマクロブロックという。マクロブロックごとに処理のモードが適応的に変更される。16 画素 × 16 ライン = 256 画素をひとまとまりとして処理を行う。H.264/AVC も既存の圧縮方式も、すべて同じサイズを使用している。
色空間	通常、画像を構成する各画素(1 画素は RGB の 3 原色で構成)の R、G、B にはそれぞれ 8 ビットが割り当てられている(1 画素 = 24 ビットとなる)。このように 1 画素が 3 種類のデータで構成されているので、これを 3 次元データとみなして、RGB「色空間」呼んでいる。	ブロック	マクロブロックの中で、それをさらに分割して、動き補償や DCT 変換を行う際の単位をブロックという。既存方式では、8 画素 × 8 ライン = 64 画素単位であった。H.264/AVC では、画像の性質によりさまざまなブロックサイズが用意され、状況に応じて最適なサイズで処理される。
Y Cb Cr	Y は輝度信号(画面の明るさを示す成分)、Cb Cr は色差信号(色の種類や強さを示す成分)。色の三原色のカラーテレビ信号 RGB(12 MHz)の色空間を Y Cb Cr(6 MHz)の色空間に変換することによって、アナログ信号を 12 MHz から 6 MHz へと帯域圧縮している。	画素精度	画素精度とは、画面を構成する 1 画素の何分の 1 の精度まで動き補償を行えるかという精度であり、より細かいほうがきめ細かい動き補償を行うことができる。画素精度には、整数精度(整数精度の画素間隔)のほか、1/2 画素精度(MPEG-2)、1/4 画素精度(H.264/AVC)などがある。
XYZ	XYZ 色空間は、デジタルシネマなどの使用を想定して規格化された色空間。		



[インターネットマガジン バックナンバーアーカイブ] ご利用上の注意

このPDFファイルは、株式会社インプレスR&D(株式会社インプレスから分割)が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

<http://i.impressRD.jp/bn>

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容(技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど)は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者(執筆者、写真の撮影者、イラストの作成者、編集部など)が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元(株式会社インプレス R&D)、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D(当時は株式会社インプレス)と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接のおよび間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先

株式会社インプレスR&D

All-in-One INTERNET magazine 編集部

im-info@impress.co.jp