2005－AVM－48（10）
2005／3／11

MPEG におけるスケーラブルビデオ符号化の標準化動向

木本崇博

NEC メディア情報研究所

概要:動画データの解像度･フレームレート･ビットレートの変換をデータの一部の削除だけで実現するスケーラブルビデオ符号化(SVC)は、発達するマルチメディア環境における重要技術と考えられている。MPEG と VCEG は、H.264/AVC の拡張ツールとして SVC の標準化を進めている。SVC で加えられたツールとして、時間方向にサブバンド分割を行う Motion Compensated Temporal Filtering と Bitplane 符号化がある。現行の SVC は、H.264 と比較して演算量と所要メモリが増加するが、同等の符号化性能を持つ。MPEG では 2006 年の規格化に向け、更なる性能改善、実利用に伴う機能追加の検討が行われている。

Activity on Scalable Video Coding in MPEG

Takahiro Kimoto

Media and Information Research Laboratories, NEC Corporation

Abstract: Scalable video coding (SVC) is considered as a proponent technique for developing multimedia environment, since it enables adaptive bitsteam conversion with different resolution, frame rate, and bitrate by simple bitstream truncation. ISO/IEC MPEG and ITU-T VCEG have jointly carried out SVC standardization as an extension tool of H.264/AVC. In SVC currently defined in MPEG, Motion Compensated Temporal Filtering and Bitplane coding are newly added into an H.264-based architecture. Although SVC requires additive computational complexity and working memory, SVC attains comparable coding performance with H.264. MPEG has developed SVC in view of coding performance and new functionalities until the SVC standard is finalized in 2006.

1. はじめに

スケーラブル符号化とは、粗い情報から細かい情報へと階層的に符号化する技術であり、静止画では Interlace GIF や Progressive JPEG など広く利用されている。動画像をスケーラブルに符号化することで、データをそのまま復号すれば高品質な動画を再生でき、一部分だけを切り出せば低解像度や低フレームレート、低ビットレートの符号化データを作ることができる。

スケーラブルビデオ符号化 (Scalable Video Coding, 以下 SVC) の研究は古くから行われ、これまでも MPEG-2/4 の一部として標準規格が決められてきた。ただし、符号化性能に劣ること、規格化当時は既存の非スケーラブルな符号化方式と置き換えて大きな効果のあるサービス基盤がなかったことから使われることは稀だった。しかし近年における Internet の爆発的な成長、伝送･視聴環境の多様化に伴い、さまざまな視聴条件に適応的な動画配信を簡易に行うことが強く求められている。粗い情報を切り出すだけで種々の視聴条件に対応できる SVC は、将来の動画配信に必須の技術だといえる。

MPEG では 2001 年から SVC の標準化を行っている。2005 年 1 月、SVC は H.264/MPEG-4 AVC の拡張ツール(Amendment)として ITU-T VCEG と共同で標準化されることとなった[1]。本報告では、現時点での SVC の技術概要と、方式改善の動向を述べる。

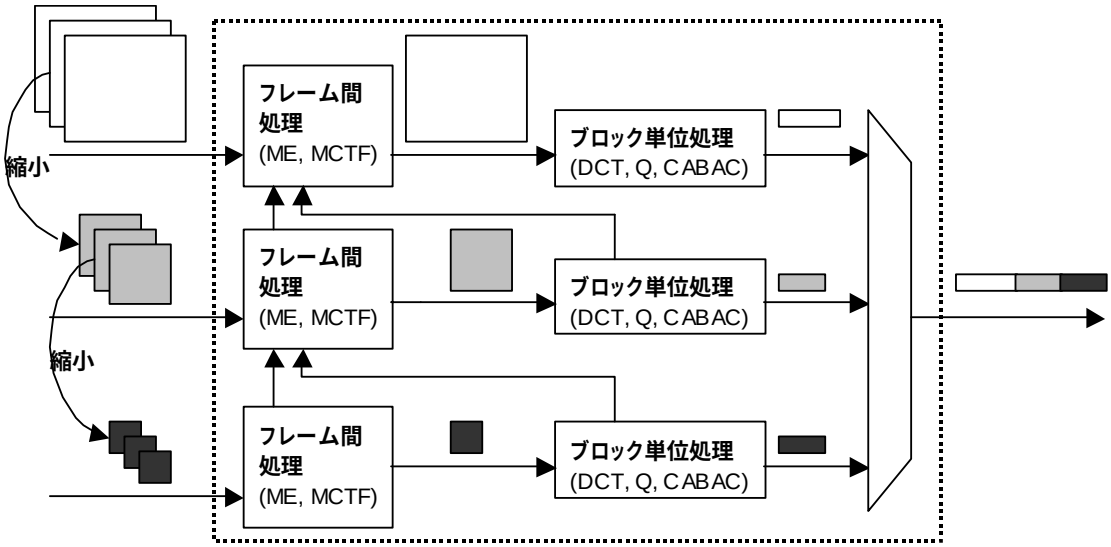

図1：SVC エンコーダの構成

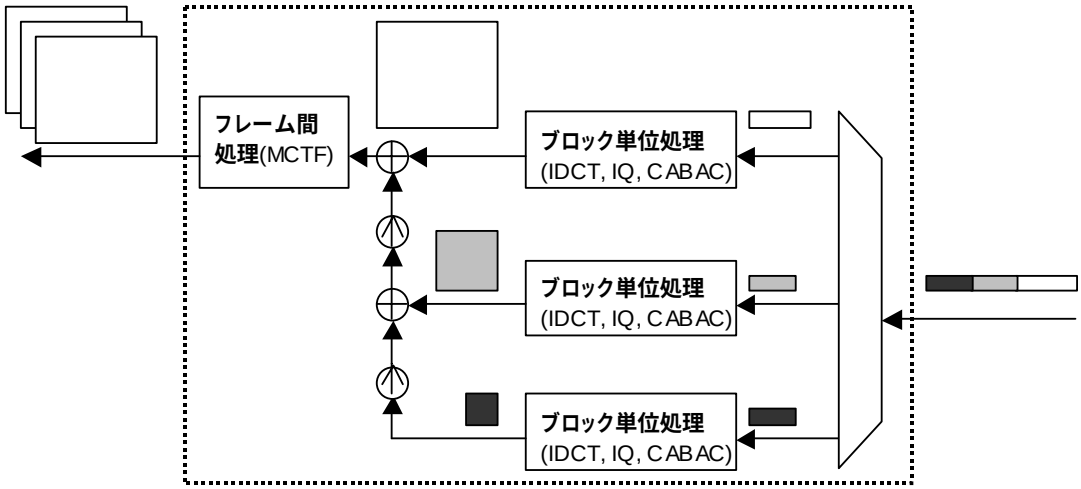

図2：SVC デコーダの構成

2. SVC の概要

2.1. 全体構成

図1および2はそれぞれSVCエンコーダ･デコーダの構成を示すブロック図である。基本的には複数の H.264 エンコーダ/デコーダを積み重ねた構成 (ピラミッド符号化)をしている[1]。エンコーダでは、解像度変換（空間スケーラビリティ）に対応するため、最大解像度の入力画像とその縮小画像信号を入力とする。各階層の入力画像について動き推定(ME)、時間方向のサブバンド分割(MCTF)からなるフレーム間処理を行う。フレームレート変換（時間スケーラビリティ）を実現するため、この処理を時間方向に再帰的に行う。同時に、下位階層の符号化情報（ローカルデコードしたテクスチャ信号、符号化モード、動きベクトル）を参照して階層間予測を

行う。その後、ブロック毎に周波数変換(DCT), 量子化(Q), エントロピー符号化(CABAC)を行う。最後に各階層の符号化データを多重化して全体の符号化データを生成する。デコーダでは、逆多重化して得られた各階層の符号化データに対し、ブロック毎に復号処理（CABAC、IQ、IDCT）を行う。その後、所望の解像度におけるサブバンドを階層間予測に基づき再構成した後、フレーム間予測処理の逆変換で復号画像を得る。

符号化時には、時間/空間解像度とビットレートの下限を設定する。この下限の階層をベースレイヤと呼ぶ。ベースレイヤでは、フレーム間予測処理・フレーム内ブロック処理ともに H.264 と同じ処理を用いる。即ちベースレイヤビットストリームは H.264 互換である。

2.2. SVC 特有のツール

2.2.1. 動き補償付時間フィルタリング（Motion Compensation Temporal Filtering, MCTF）

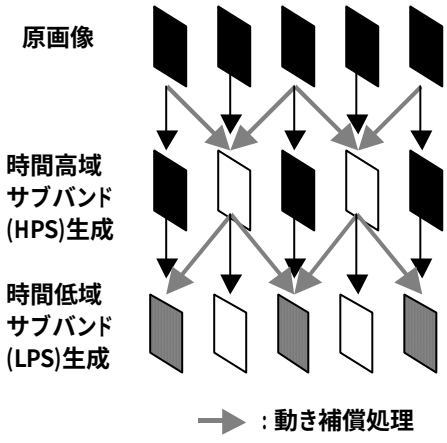

図3： MCTF

図3は MCTF の処理を表す概念図である。MCTF はフレーム間処理の一つであり、レート変換（SNR スケーラビリティ）でのドリフトによる画質劣化を低減させるための重要な技術である。MPEG-1/2/4 で行われる動き補償予測符号化では、参照フレームの復号画像を参照し、動き補償処理によって予測信号を生成した後、現フレームとの誤差信号を符号化する。一方、MCTF では、時間方向に連続する信号をサブバンド分割する。参照フレームの原画像を参照し、動き補償予測によって予測信号を生成した後、現フレームとの誤差信号、即ち時間高域サブバンド(High-Pass Subband, HPS)を生成する。さらに、予測誤差信号に動き補償処理を行い、参照フレームに足しこむことで時間低域サブバンド(Low-Pass Subband, LPS)を得る。時間方向の変換が直交になることで、各フレームで量子化誤差のパワーが保持される。これによりレート変換での段階的量子化によるドリフトを抑制することができる。1GOP 内で図4に示すように L1, L2 と再帰的に MCTF を行うことで、時間スケーラビリティを実現する。図で L1,2,3 は LPS、H1,2,3 は HPS を示す。

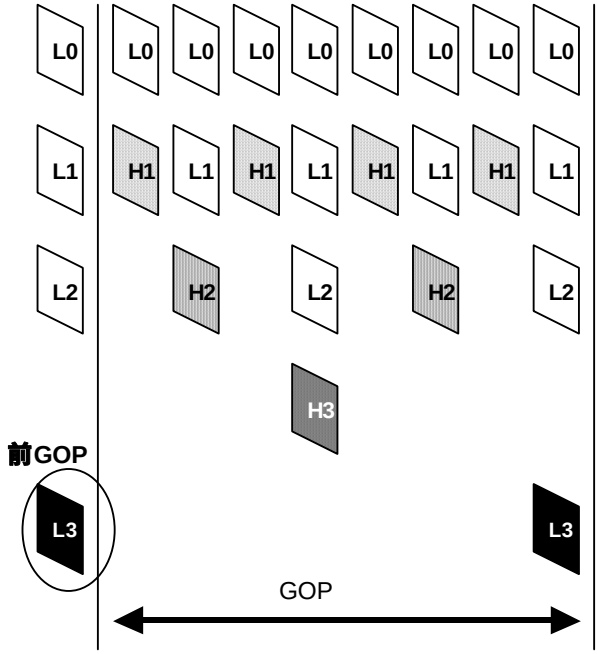

図4：MCTF の階層化

2.2.2 Bitplane 符号化

数ビット単位の細かいレート変換(Fine Grain Scalability, FGS) を実現するため、SVC では JPEG2000 と同様に Bitplane 符号化を用いる。

図 5 に一次元配列での Bitplane 符号化の例を示す。ある解像度の最低ビットレートでの量子化ステップをΔとする。まず、そのステップで量子化した係数を符号化する。図では（左から）1, 3 番目の係数が非ゼロであり、正負符号とレベルが符号化される。次に量子化誤差を量子化ステップΔ/2 で符号化する。図では 1, 3 番目に加えて 2 番目の係数が新たに非ゼロ係数として正負符号が符号化される。最大レベルは常に 1 である。以下、量子化ステップを 1/2 にしながら誤差を段階的に符号化する。最低ビットレートでは H.264 と互換性があるように符号化される。

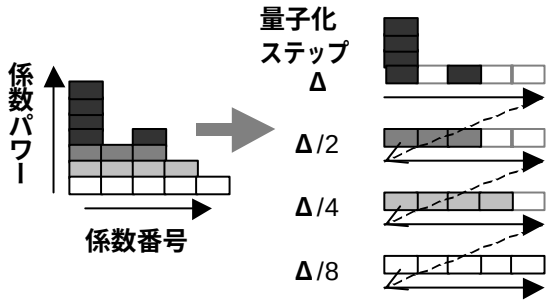

図5：Bitplane 符号化

2.2.3. 階層間予測

SVC では、符号化効率向上のため、階層間の相関を利用する。テクスチャの符号化では下位階層のイントラ信号および残差信号を、符号化モード・動きベクトルの符号化では下位階層の情報を予測に用いる。

図6にテクスチャにおける階層間予測を示す。あるブロックに対し、同じ位置にある下位階層のブロックがイントラ符号化されている時、当該ブロックの符号化モードとして下位階層との差分を符号化することを選択できる(Intra_BL モード)。また、あるブロックにおいて、フレーム間予測による残差信号が符号化されるとき、残差予測フラグを付加情報として符号化する。残差予測フラグが1の時、同じ位置にある下位階層の残差信号との差分を符号化する。下位階層の解像度が違う時、拡大してからこれらの処理を行う。なお、Intra_BL モードは HPS でも利用できる。

図7に符号化モード・動きベクトルの階層間予測を示す。マクロブロック毎に、階層間予測を制御する付加情報である base_mode_flag, base_mode_refinement_flag を併せて符号化する。base_mode_flag が1の時、下位階層の符号化モードと動きベクトルをそのまま上位階層に引き継ぐ。下位階層の解像度が違う場合、これらを解像度比率に従って伸張する。また、base_mode_refinement_flag が1の時には、同様に伸張した後、サブピクセルの分だけ動きベクトルの補正分を符号化する。

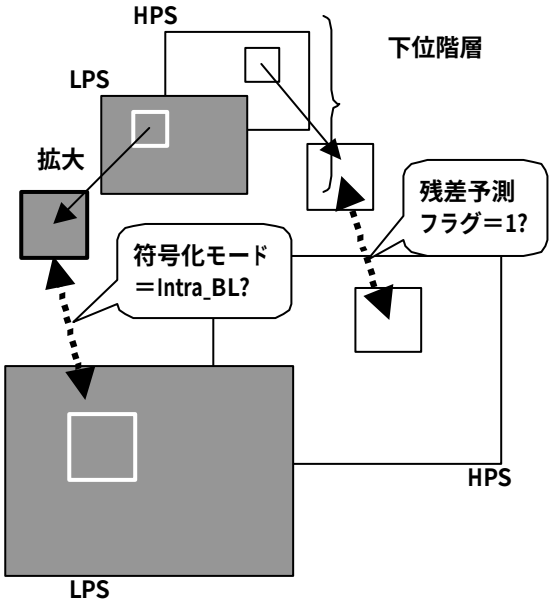

図6：テクスチャにおける階層間予測

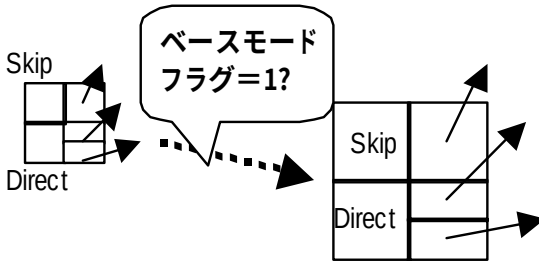

図7：符号化モード・動きベクトルにおける階層間予測

2.3. 処理コストの評価

H.264 と比較した SVC デコーダにおけるコスト（演算量、所要メモリ）を評価する。

- フレーム内ブロック単位処理

CABAC が Bitplane 対応となるが、発生符号量が同じであれば演算量の差は殆どない。また、各階層で IQ と IDCT を行う必要があるが、演算量、所要メモリとも増加量はごくわずかである。

- フレーム間予測処理

H.264 のフレーム間予測では動き補償処理をフレーム毎に 1 回行うのに対し、MCTF では 2 回になる。例えば、デコーダに要する演算量の 1/3 が動き補償処理だとすればSVC デコーダの総演算量は H.264 に比べて 30%増加となる。また MCTF では、図 4 に示すように、LPS を生成するのに過去の HPS も保持しておく必要がある。そのため、所要メモリは 2 倍近くとなる。

2.4. 方式改善の動向

H.264 コーデックを積み重ねる旧来からの枠組みを用いる SVC の現行方式に対し、標準化作業で種々の方式改善が検討されている[2]。

符号化性能向上のため、階層間相関をさらに利用する方法が多く検討されている。筆者らは、3D ウェーブレット技術に対して既に提案した、下位階層の復号画像に上位階層の動きベクトルを用い動き補償を行う方式 [3] を現行 SVC 方式に応用し、符号化性能の向上を確認している。また、Samsung, LG らは階層間相関を用いた時のブロック歪み除去フィルタ調節や、動きベクトルの符号化方式についてそれぞれ提案している。

符号化データ生成処理についても SVC 特有の改善検討がなされている。Nokia は Bitplane 符号化の効率化について多く提案している。また、France Telecom は階層間の符号割り当て配分について検討を行っている。

実利用時に要求される機能についても検討がなされている。Panasonic は、現行の MCTF では低遅延配信を実現できないことから、多くの提案を行っている。また、Thomson は、HD から SD への変換などサイズ比が 2 倍単位でない解像度変換の必要性について報告している。

3. 符号化性能

[4]に報告されている実験結果を示し、SVC の符号化効率を評価する。4CIF サイズ(704x576 ピクセル)、60fps の画像”City”, “Crew”に対し、表 1 に示すように時間空間方向に 3 階層、計 6 階層からなるビットストリームを生成し、それぞれについて復号画像の PSNR を比較する。

表 1 : ビットストリームの階層

解像度		ビットレート [kbps]	
空間	時間	“City”	“Crew”
QCIF	15fps	64	96
QCIF	15fps	128	192
CIF	30fps	256	384
CIF	30fps	512	750
4CIF	30fps	1024	1500
4CIF	60fps	2048	3000

SVC では、1 度の符号化で階層化された 1 本のビットストリームを生成する。得られたビットストリームから、各階層に相当するサブビットストリームを抽出し、それぞれを復号する。GOP サイズは City で 32, Crew で 16 である。FGS に対応する(w/FGS)場合と対応しない(wo/FGS)場合、双方について評価を行う。一方、比較対象の H.264 では表 1 に示す条件それぞれに独立に符号化･復号する。I ピクチャ間隔はいずれも 32、P ピクチャ間の B ピクチャの枚数は 2 とし、参照ソフト JM8.1 を用いた。

図 8、9 にそれぞれの画像における各階層での輝度成分の平均 PSNR を記す。動きの緩やかな”City”では、4CIF 解像度において H.264 に比べ、0.6dB 程度画質が向上している。逆に、カメラのフラッシュが入りフレーム間の相関の低い”Crew”では、4CIF 解像度において FGS なしの場合で最大 0.5dB 程度画質が低下している。また、FGS に対応した場合、”City”では最大 0.7dB 程度、”Crew”では 0.3dB 程度画質が低下している。

4．標準化作業の動向

MPEG の SVC 標準化では最初、多くの提案方式が時間/空間方向にウェーブレット変換を行う3Dウェーブレットを用いていた。しかし、2004年10 月のパルマ会合で行われた主観評価実験で、HHI (独) の提案した H.264 をベースとした方式が全ての符号化条件で優れていたため、基本方式に選ばれた。またそれまで MPEG-21 の1パートだった SVC は、H.264/MPEG-4 Part 10 AVC の拡張ツールとして新たに位置づけられた。今後は、2005 年 10 月に Committee Draft、2006 年 7 月に Final Draft International Standard の策定を目標に作業が進められる。

一方、3D ウェーブレットについては、現在でも MPEG で性能評価が続けられている。3D ウェーブレットは現在の標準 SVC 方式と比較して、細部の再現性に優れているが、動きの激しい画像など符号化の困難な場合に歪みが目立ちやすい、動きベクトルとテクスチャの符号量割り当てが難しい、などの問題がある。これは、3D ウェーブレットは現在研究段階にあり、H.264 のように主観画質向上のための細かいノウハウがまだ蓄積されていないためだと言える。標準 SVC 技術が 3D ウェーブレットに置き換わることはないものの、次世代符号化技術として多くの団体が性能改善の検討を続けている。

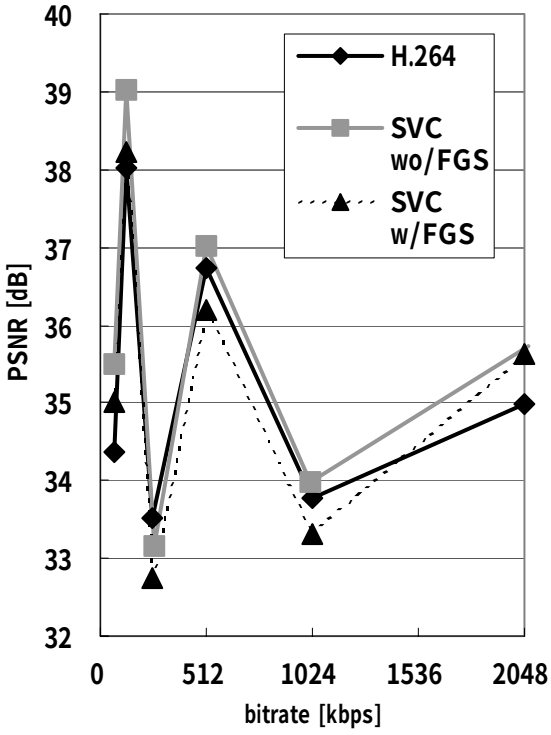

図8：平均 PSNR (画像”City”)

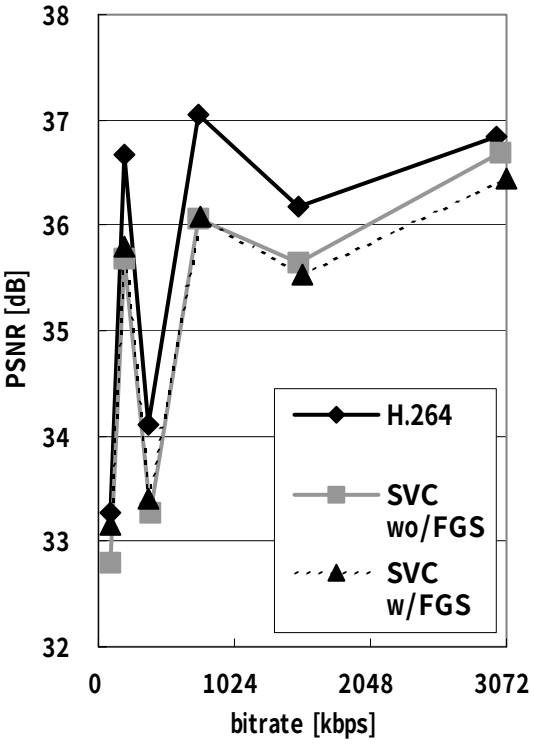

図 9：平均 PSNR (画像”Crew”)

参考文献

[1] Joint Video Team of ISO/IEC MPEG & ITU-T VCEG, "Working Draft 1.0 of 14496-10:200x/AMD1 Scalable Video Coding", ISO/IEC JTC1/SC29/WG11 N6901, Jan. 2005

[2] “Description of Core Experiments for SVC”, ISO/IEC JTC1/SC29/WG11 N6898, Jan. 2005

[3] 木本ら,「多重階層時間フィルタリングを用いる三次元ウェーブレット符号化」, PCSJ2004, P-2.13

[4] H. Schwarz, et.al. “Technical description of the HHI proposal for SVC CE1”, ISO/IEC JTC1/SC29/WG11 M11244, Palma, Oct. 2004