2007－AVM－58 (3)
2007／9／25

空間スケーラブル符号化における低解像度信号を利用したプレフィルタ設計に関する考察

早瀬和也, 坂東幸浩, 高村誠之, 上倉一人, 八島由幸

日本電信電話株式会社　NTTサイバースペース研究所

あらまし 空間解像度の多様化に柔軟に対応しうるH.264/AVCのスケーラブル拡張JSVCが標準方式として規格化された。今後は、JSVCエンコーダの符号化性能向上が重要となる。映像符号化の性能向上手法として、プレフィルタが使用される。視覚的に重要でない高周波成分をフィルタで落とすことで、主観画質を保持したまま符号量を削減できる。しかし、従来のプレフィルタはシングルレイヤ符号化への適用を前提としており、マルチレイヤ構造の符号化にこれをそのまま適用しても、レイヤ間の冗長性除去には結びつかない。そこで本稿では、レイヤ間の冗長性除去を積極的に狙い、拡張レイヤにおけるレイヤ間予測を利用したプレフィルタを設計した。提案手法により、主観画質同等の条件下で平均9.1%の符号量削減を実現した。

キーワード スケーラブル符号化、レイヤ間予測、プレフィルタ

A study of designing pre-filter using low resolution signals for spatial scalable video coding

Kazuya HAYASE, Yukihiro BANDOH, Seishi TAKAMURA,
Kazuto KAMIKURA, Yoshiyuki YASHIMA

NTT Cyber Space Laboratories, NTT Corporation

Abstract: JSVC, which is the scalable extention of H.264/AVC, was standardized. It will be an important issue to develop efficient JSVC encoders. Pre-filters are often utilized to enhance coding efficiency. Pre-filters remove high frequency energy, which is little important for human visual system, to reduce coding bits without degrading image quality. However, previous pre-filters are designed for single-layer coding. Even if a previous pre-filter is applied to multi-layer coding, such as JSVC, inter-layer redundancy cannot be reduced. We proposed an enhancement layer pre-filter using inter-layer prediction to reduce inter-layer redundancy. As a result, the proposed pre-filter could reduce coding bits by 9.1%.

Key words: scalable video coding, inter-layer prediction, pre-filter

1　はじめに

今や映像コンテンツの閲覧媒体は、携帯端末、PC、テレビ、大型スクリーンと多岐に及び、その空間解像度は多様化を続けている。近年、このような空間解像度の多様性に柔軟に対応しうるスケーラブル符号化方式への期待が高まる。ITU-TとISOの合弁標準化団体JVTは、2007年7月に標準スケーラブル符号化方式Joint Scalable Video Coding（以下JSVC）[1]を策定した。この方式は、高圧縮率を実現したH.264/AVCのスケーラブル拡張として位置づけられており、高圧縮率と高スケーラビリティの両立を目指している。今後は、JSVCエンコーダの符号化性能向上が重要となる。

映像符号化における符号化効率向上の一手段とし

て、プレフィルタが使用される。視覚的に重要でない空間高周波成分をローパスフィルタを用いてあらかじめ落とすことで、主観画質を保持したまま符号量を削減できる。JSVC の符号化処理は、空間スケーラビリティをサポートするために、所望の空間解像度ごとにレイヤを成すレイヤ構造をとっており、この各レイヤごとにプレフィルタを適用することで符号化性能の向上が期待できる。

しかし、前述した従来のプレフィルタは、非スケーラブルのシングルレイヤ符号化への適用を前提としている。したがって、JSVC の各レイヤに対して従来のプレフィルタをそのまま適用しても、レイヤ間の冗長性の除去には直接結びつかない。JSVC のみならず、基本レイヤと拡張レイヤを持つマルチレイヤ構造の空間スケーラブル符号化への適用には、改善の余地を残す。

そこで本稿では、拡張レイヤにおけるレイヤ間予測を利用したプレフィルタの設計を行う。提案するプレフィルタは、レイヤ間の情報の冗長性除去を積極的に狙う。主観画質を保持したまま、符号化前段においてあらかじめ不必要なレイヤ間の情報を削減することにより、符号化性能の改善を図る。本稿では、プレフィルタの設計手順を述べ、その性能検証を主観評価実験を通して行う。

2 JSVM の符号化処理概要

JSVC の符号化参照ソフトウェアは、Joint Scalable Video Model（以下 JSVM）[2] と呼ばれる。JSVM における空間スケーラビリティの実現構造を図 1 に示す。図 1 は、2 つの解像度のスケーラビリティを持つときの例である。JSVC では、基本レイヤと拡張レイヤから構成されるマルチレイヤ構造により空間スケーラビリティを実現している。入力した原信号が拡張レイヤへの入力信号、その縮小信号（通常は縦横半分）が基本レイヤへの入力信号となり、基本レイヤ、拡張レイヤの順に符号化を行う。基本レイヤでは、H.264/AVC に準拠した符号化処理を行うのに対し、拡張レイヤでは、H.264/AVC に備えられている画面内予測および画面間予測に加えて、基本レイヤの符号化済み情報から予測するレイヤ間予測を行い、マクロブロック単位で 3 つの予測の中で最も効率が良いものを適宜選択する。

このレイヤ間予測は、テクスチャ予測と動き補償情報予測の 2 つに大別される。テクスチャ予測とは、基本レイヤの符号化ストリームを符号化処理中に一時復号し、それを拡張レイヤの解像度までアップサンプルした信号を拡張レイヤにおける予測参照信号とする方式である。動き補償情報予測とは、基本レイヤにおいて導出された動き補償情報（動きベクトルや参照フレーム番号など）を拡張レイヤに流用する方式である。

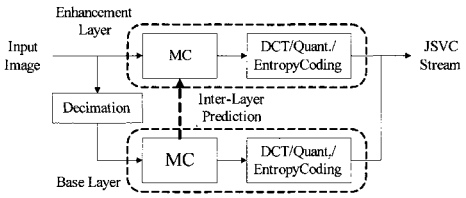

図 1: JSVM の符号化処理図

3 提案プレフィルタ

3.1 着眼点

前述したように、レイヤ間予測方式の一つに、符号化済みの基本レイヤの復号信号をアップサンプルし、それを予測参照信号とするテクスチャ予測がある。このテクスチャ予測信号は、サンプリング前の信号の帯域以上の高周波成分は含んでいないため、原信号に空間ローパスフィルタをかけた状態と等価とみなせる。仮に、プレフィルタ処理により、拡張レイヤの符号化処理への入力信号が、アップサンプリングにより得られるテクスチャ予測信号に置換されたと仮定する。すると、拡張レイヤの符号化処理において、テクスチャ予測の予測残差信号が必ずゼロとなるため、符号量は大幅に削減される。しかし、拡張レイヤ内の全画素に対して前述のプレフィルタ処理が適用されるとアップサンプル信号の高周波成分の欠落によるボケが顕在化し、主観画質は損なわれる。

そこで、本提案方式では、画質劣化が見えやすい部分には拡張レイヤの信号を適用し、画質劣化が見えにくい部分にはテクスチャ予測信号を、拡張レイヤの符号化入力信号として扱う。つまり、画質劣化が見えにくい部分についてのみ、拡張レイヤの原信号をテクスチャ予測信号で置換する。本置換作業をプレフィルタ処理とみなす。

3.2 フィルタリング手順

本プレフィルタは、基本レイヤの符号化が既に済んでいる前提のもと行われる。はじめに、拡張レイヤの原信号とテクスチャ予測信号を入力として、マ

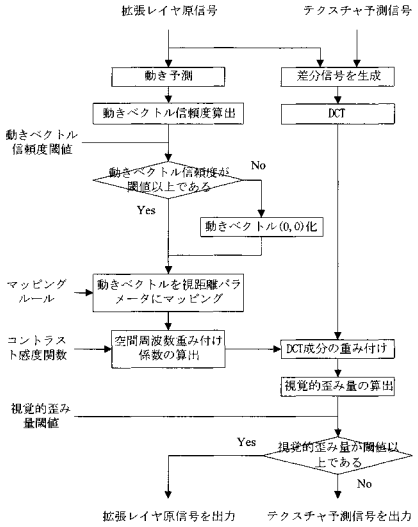

図 2: プレフィルタ処理の流れ

クロブロックごとに両信号の画質比較を行う。画質の比較尺度には、単純な画素値の二乗誤差ではなく、視覚的歪み量を利用する [3]。この視覚的歪み量は、人間の視覚が持つ時空間のコントラスト感度に基づいている。視覚的歪み量の算出モデルは、次節において詳説する。まず、拡張レイヤの原信号と補間信号との間の視覚的歪み量 D を計算する。この視覚的歪み量は、劣化の目立ちやすさを表す尺度であり、値が小さいほど目立ちにくいということを示す。算出手順は次節で詳説する。この視覚的歪み量が閾値 D_{th} 以下であれば、原信号をテクスチャ予測信号で置換する。$D > D_{th}$ であれば、置換を行わず、原信号をそのまま出力する。この閾値は、劣化が知覚されない、もしくは、気にならない視覚的歪み量の許容値であり、外部より与える。この視覚的歪み量に基づいた適応制御により、視覚劣化が目立ちにくい領域のみテクスチャ予測の残差信号をゼロにでき、主観画質を損なわずに符号量を削減できる。

マクロブロックにおける本プレフィルタの全体の処理フローを図 2 に示す。

3.3 視覚的歪み量の算出

本プレフィルタに用いた視覚的歪み量算出モデルは、人間の視覚系の持つ時空間のコントラスト感度に基づいている [4]。人間の視覚系は、空間高周波成分に対するコントラスト感度が空間低周波成分に対して相対的に低い。これは、空間高周波成分の欠落による画質劣化が知覚されにくいと解釈できる。その傾向は、時間周波数が高くなるにつれて、より顕著になる。したがって、時間高周波成分を多く含む動きの早いシーンでは、空間高周波成分の欠落による画質劣化は静止シーンにおけるそれよりも知覚されにくい。これらの性質を考慮して作成された歪みモデルは、通常の画素値の二乗誤差よりも、主観画質を良好に反映した定量的劣化指標となりうる。

$M \times M$ ブロックの視覚的歪み量 D は、次のように求める。

$$D = \frac{1}{M^2} \sum_{u=1}^{M} \sum_{v=1}^{M} F(u,v) \cdot w(u,v) \qquad (1)$$

ここで、$F(u,v)$ は 2 つの対象信号の当該ブロックにおける差分信号に対する空間周波数 (u,v) のフーリエ係数である。$w(u,v)$ はそのフーリエ係数に対する重み係数であり、空間周波数に対するコントラスト感度を指す。この値は、文献 [5] の空間コントラスト感度関数に基づいて、次式のように算出する。

$$w(\eta) = (0.2 + 0.45\eta)\exp(-0.18\eta) \qquad (2)$$

ここで、η と (u,v) は、$\eta = \frac{\sqrt{u^2+v^2}}{\theta}$ [cycle/degree] の関係にある。この θ は、縦幅 h の画像を視距離 ah で観測する場合の MxM ブロックの視野角であり、次のように与えられる。

$$\theta = \frac{360}{\pi} \arctan\left(\frac{M}{2ah}\right) \text{ [degree]} \qquad (3)$$

この視距離パラメータ h は、通常は実験環境に応じて設定される。視距離が遠いとコントラスト感度は低く、近いとコントラスト感度は高い。ここでは、この視距離パラメータを時間コントラスト感度の制御パラメータとして用いる。動きが早い領域ではコントラスト感度が低いため視距離パラメータを大きくし、静止領域では小さくする。領域の動き量を示す指標としては、その領域の前フレームからの動きベクトルのユークリッドノルム n の値を適用する。動きベクトルのノルムと視距離パラメータのマッピングルール $(n \to h)$ は、経験的に作成する。

3.4 動きベクトル信頼度の算出

動きベクトルのノルムと視距離パラメータのマッピングにあたり、動きベクトルの信頼度という要素を考慮する。ブロックマッチングにより算出される

動きベクトルは、単色領域や繰り返し模様の領域において、真の動きと乖離する可能性が高い。したがって、真の動き量が小さいにも関わらず、大きいノルムを持つ動きベクトルが推定されてコントラスト感度が低いとみなされた場合には、画質劣化が見られる可能性が高くなる。そこで、動きベクトルと真の動きとの相似度を示す動きベクトル信頼度 R を算出する。この信頼度 R が閾値 R_{th} より低いときには、視距離パラメータとマッピングさせる動きベクトルを強制的に $(0,0)$ とする。

時刻 t のブロック B に対して時刻 $t-1$ フレームにおいて動き予測をし、決定された予測参照ブロック B' からの動きベクトルを $\mathbf{v}_{(B'_{t-1}\to B_t)}$ とする。この動きベクトルの信頼度を次のように求める。時刻 $t-1$ のブロック B' に対して時刻 t フレームにおいて動き予測をし、決定された予測参照ブロック B'' からの動きベクトル $\mathbf{v}_{(B''_t\to B'_{t-1})}$ を求める。真の動きは1つしか存在しないため、動きベクトル $\mathbf{v}_{(B'_{t-1}\to B_t)}$ が真の動きを正確に反映しているならば、それは $-\mathbf{v}_{(B''_t\to B'_{t-1})}$ と一致するはずである。したがって、この両ベクトルの相似度を動きベクトルの信頼度としてみなせる。本研究では、次式のように両ベクトルの内積値を信頼度として用いた。

$$R = -\mathbf{v}_{(B'_{t-1}\to B_t)} \cdot \mathbf{v}_{(B''_t\to B'_{t-1})} \tag{4}$$

4 視覚的歪み量の閾値の設定

4.1 適正閾値推定の必要性

視覚的歪み量の閾値 D_{th} は、大きくすれば符号量の削減が多く見込めるものの、復号信号の画質はプレフィルタなしの場合よりも劣化する可能性が高まる。反面、小さくすれば画質劣化は避けられるが符号量の削減は小さくなる。したがって、復号信号の画質がプレフィルタありとなしの場合で主観的に差が見られない限界の閾値が最も望ましい、と判断できる。通常、この適正閾値は、閾値を徐々に変化させていき、随時プレフィルタなしの場合の復号信号と主観画質の比較を行っていくことで同定するが、この作業は非常に時間と人手がかかり非効率である。かといって、この適正閾値は映像や閲覧環境によって異なってくるため、一意にモデル化することは難しい。

そこで、本研究では、符号化性能をやや犠牲にし、実用面を重視した主観画質の損失を確実に避けることができる安全な閾値の同定を、主観評価実験を通して試みる。

表 1: 主観評価実験条件

シーケンス	City、Crew、Harbour、Soccer
解像度	（基本）CIF（拡張）4CIF
フレーム数	300 [frames]（30 [fps]）
ベース画像	プレフィルタなしの場合の復号信号
比較画像	ベース画像、閾値 4/6/8 の復号信号
視距離	3H
評価者	（専門家）4 人（非専門家）4 人

表 2: 評価基準

5	劣化が分からない
4	劣化が分かるが気にならない
3	劣化が気になるが邪魔にならない
2	劣化が邪魔になる
1	劣化が非常に邪魔になる

4.2 閾値設定のための主観評価実験

実験条件を表 1 に示す。評価実験には、ベース画像と比較画像を並べて同時に投影し、ベース画像に対する比較画像の劣化度合いを測る SDSCE（同時二重刺激連続評価）法を用いた。プレフィルタなしの場合の復号信号をベース画像に、視覚的歪み量の閾値 D_{th} を 4、6、8 と設定してプレフィルタを適用した場合の復号信号を比較画像とし、ベース画像に対する画質劣化の評価を行った。また、ベース画像同士の比較も、評価者の評価の信頼性を確認するために行った。評価者は、符号化の専門家 4 人と非専門家 4 人の計 8 人とした。主観評価の評価基準を表 2 に示す。実験映像は 10 秒であり、評価者は 1 つの実験映像を 3 秒間の灰色画像を挟んで 2 回見、その後再度 5 秒間の灰色画像を挟んで次の実験映像に移る。評価の記入は、2 回目の実験映像の表示中と 5 秒間の灰色画像表示中の計 15 秒の間に行う。

主観評価実験の結果を図 3 に示す。結果は、紙面の都合上、量子化パラメータが 18 の場合の City と Soccer における評価結果のみ載せる。なお、量子化パラメータを 24、30 と変えた場合も図 3 とほぼ同様の傾向を示すことを確認している。プロットは、評価点の平均値であり、負の方向に伸びるエラーバーは評価点の標準偏差を示している。評価点 4 を劣化知覚の境界線とみなすと、どの映像についても閾値 4 であれば、気になるほどの劣化は知覚されないと判断できる。Crew、Harbour でも同様のことが言えることを確認している。City の復号信号の例を図 4 に示す。閾値 6 になると、ブロックノイズや細部の

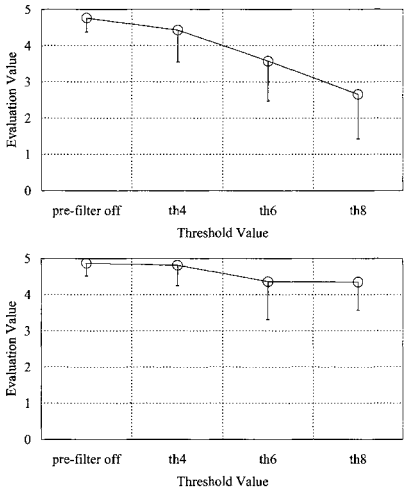

図 3: 主観評価結果（上：City 下：Soccer）

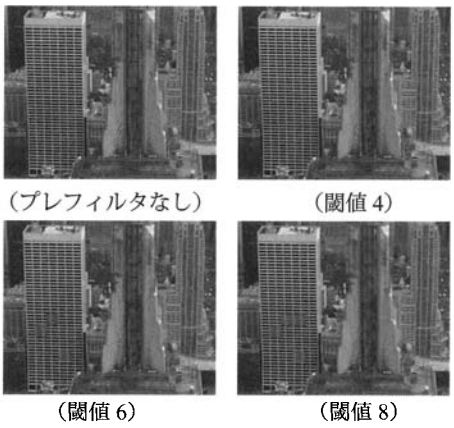

図 4: 評価画像例（City）

情報の欠損が顕在化していることが分かる。

5 実験

5.1 実験条件

提案手法を JSVM ver.8.0 に実装した。そして、提案手法と JSVM の性能比較を行った。

符号化実験条件を表 3 に示す。JSVC 標準画像の 4 つを実験映像として用いた。解像度は、基本・拡張レイヤがそれぞれ、QCIF・CIF（ケース 1）、CIF・4CIF（ケース 2）の 2 ケース行った。ピクチャ構造は IPPP

表 3: 符号化実験条件

エンコーダ	JSVM ver.8.0
シーケンス	City, Crew, Harbour, Soccer
解像度（基本/拡張）	ケース 1　QCIF/CIF ケース 2　CIF/4CIF
フレーム数	300 [frames]（30 [fps]）
GOP	IPPP
QP	18, 24, 30
視覚的歪み量閾値 D_{th}	4 [per pixel]
MV 信頼度閾値 R_{th}	0
n と h の対応付け	$h = 2H$　if $n \leq 5$ $h = 5H$　else

とし、各映像 300 フレームの符号化を行った。量子化パラメータ（QP）は、18、24、30 の 3 パターン試し、両レイヤで同じ値を使用した。視覚的歪み量閾値 D_{th} は、4.2 節の結果を踏まえ、主観画質の劣化を確実に回避できる値の 4 とした。動きベクトル信頼度閾値 R_{th} は 0 とした。

5.2 結果と考察

符号量の削減結果を表 4 に示す。提案手法は、ケース 1 で JSVM 比平均 5.7% の符号量削減、ケース 2 で平均 9.1% の符号量削減を実現している。また、その削減効果は、量子化パラメータが小さくなるにつれて、つまり高レートになるにつれてが大きいことが分かる。一般的に 1 つのブロックに発生する予測残差信号のエネルギーは、高レートになるほど大きくなる。プレフィルタによって置換されたブロックは、必ずテクスチャ予測による予測残差信号のエネルギーがゼロになるため、その効果が高レートになるほど大きくなる。

この削減効果の大きさには映像依存性がある。Crew や Soccer のような動きが大きく空間高周波成分が少ない映像では、置換が行われやすいために削減効果が大きい。反対に、City や Harbour は空間高周波成分が多いため、他シーケンスより効果が小さい。一部符号量が増加しているのは、置換による符号量減少分より、置換によって動き予測の参照先が変更されて動き予測性能が悪化したことによる増加分が上回ったためと考えられる。

上記の傾向を示す理由は、図 5 に示す解像度ケース 1 における City と Soccer のピクチャタイプ別符号量削減率の結果からも裏付けられる。提案手法は、テクスチャ予測の効果の増大が見込まれるアプローチであるため、I ピクチャにおける符号量削減効果が最も大きい。図 5 からもその傾向が見てとれる。一方、

表 4: 符号量 [kbps]（括弧内は削減率 [%]）

（ケース 1：(基本) QCIF（拡張）CIF）

		City	Crew	Harbour	Soccer
QP 18	off	3289.2	4197.6	6138.7	3329.2
	on	3302.7 (+0.41)	2860.7 (-31.8)	6131.5 (-0.12)	2873.5 (-13.7)
QP 24	off	1131.5	1872.5	3185.0	1500.9
	on	1151.8 (+1.79)	1550.3 (-17.2)	3185.6 (+0.02)	1435.7 (-4.34)
QP 30	off	374.97	813.44	1293.2	671.62
	on	377.94 (+0.79)	783.23 (-3.71)	1293.7 (+0.04)	669.59 (-0.30)

（ケース 2：(基本) CIF（拡張）4CIF）

		City	Crew	Harbour	Soccer
QP 18	off	19402	17025	23622	15244
	on	18283 (-5.77)	9797.5 (-42.5)	23142 (-2.03)	11929 (-21.7)
QP 24	off	6884.2	6655.7	11552	5784.7
	on	7027.5 (+2.08)	4951.7 (-25.6)	11520 (-0.28)	5464.5 (-5.54)
QP 30	off	1570.3	2343.3	4564.7	2080.8
	on	1592.5 (-1.41)	2189.7 (-6.55)	4569 (-0.09)	2072.6 (-0.39)

P ピクチャでは、映像の動き予測への寄与が高いほど、動き予測参照先の劣化が悪影響を及ぼす。Soccer は I、P ピクチャともに、符号量が削減されている。しかし、City は、I ピクチャでは符号量が削減されているものの、P ピクチャではわずかに符号量が増大してしまう。映像全体では P ピクチャの枚数が圧倒的に多いことから、全体の符号量は増加してしまう。

6 まとめ

本研究では、レイヤ間予測による削減効果を増大させるために、レイヤ間予測の一つであるテクスチャ予測を利用した拡張レイヤのプレフィルタを提案した。提案手法により、JSVC 参照ソフトウェアと比較して、主観画質同等の条件のもと、平均 9.1% の符号量削減を実現できた。今後は、フレームごとのフィルタの on/off 適応制御を行い、動き予測性能の劣化による符号化性能低下の回避を図る。

参考文献

[1] J. Reichel, M. Wien and H. Schwarz: "Working Draft 4 of ISO/IEC 14496–10:200x Amd.3 Scalable Video Coding," ISO/IEC JTC1/SC29/WG11, N7555, 2005.

[2] J. Reichel, H.Schwarz and M. Wien: "Joint Scalable Video Model JSVM–8.0," ISO/IEC JTC1/SC29/WG11 and ITU–T SG16 Q.6, JVT–U202, 2006.

[3] 坂東幸浩, 高村誠之, 上倉一人, 八島由幸: "主観画質を考慮した H.264/AVC におけるモード選択方法の検討," オーディオビジュアル複合処理研究会（AVM）, Sep. 2006.

[4] D. H. Kelly: "Motion and vision. II. Stabilized spatio-temporal threshold surface," J. Opt. Soc. Am. 69(10), pp. 1340–1349, Oct. 1979.

[5] N. B. Nill: "A visual model weighted cosine transform for image compression and quality assessment," IEEE Trans. Commun., Vol. COM–33, No. 12, pp.551–557, June 1985.

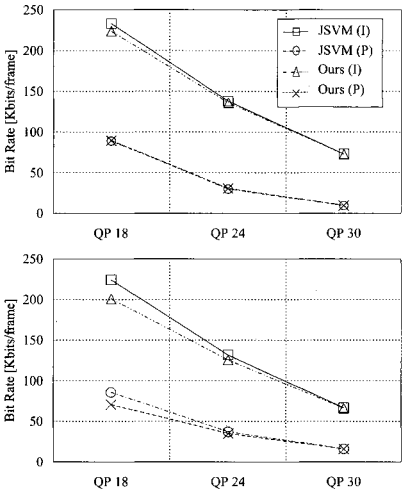

図 5: ピクチャ別符号量削減率（上：City 下：Soccer）