

LI-001

# 空間スケーラブル符号化におけるレイヤ間相関を利用した重み付き予測方式 Weighted Prediction of Spatial Scalable Video Coding using Inter-Layer Correlation

早瀬 和也, 坂東 幸浩, 高村 誠之, 上倉 一人, 八島 由幸

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

## 1 はじめに

近年のネットワーク環境や表示環境の多様化を背景に、H.264/AVC のスケーラブル拡張方式の JSVC が策定された [1]。この JSVC の処理系は、空間スケーラビリティ実現のために、所望の解像度ごとに H.264/AVC を適用して符号化情報を多重化するレイヤ構造をとる。

JSVC において継承される H.264/AVC の特徴的な符号化処理ツールの一つに、フェードのような輝度が時間的に変化する画像に対して、動き予測の性能低下を低減する重み付き予測 (WP: Weighted Prediction) がある [2]。フレーム間の輝度変化を推定し、動き予測参照信号を補正する。この輝度変化の推定情報は、WP パラメータとして算出される。H.264/AVC の重み付き予測では、WP パラメータを符号化情報として伝送する Explicit モードと、伝送せずに復号信号から構築する Implicit モードの 2 つのモードが用意されている。Explicit モードは高精度の重み付き予測を行えるが、WP パラメータを伝送するため符号量が増大する。これに対し、Implicit モードは WP パラメータの符号量を必要としないが、時間方向の内挿/外挿により WP パラメータを算出するため、重み付き予測性能の精度は低下する。

JSVC の重み付き予測の機構は、H.264/AVC のそれを各レイヤに適用したものになっている。したがって、JSVC の重み付き予測 Implicit モードも、WP パラメータは一つのレイヤに閉じた時間方向の輝度変化情報のみに基づいて算出される。しかし、JSVC では、上位レイヤにおいて重み付き予測を行う場合には、符号化済みの下位レイヤの輝度情報を使用することができるため、WP パラメータの推定精度に改善の余地を残す。

そこで、本研究では、マルチレイヤ構造をとる JSVC の特徴に注目し、レイヤ相関を利用した上位レイヤの重み付き予測の改良を試みる。WP パラメータの算出に直下レイヤの輝度情報を利用することで、Implicit モードにおける WP パラメータの推定精度を向上させる。

## 2 JSVC における重み付き予測とその問題点

図 1 は、2 つの空間スケーラビリティを実現する JSVC のブロック図である。所望の解像度の画像を原画像から

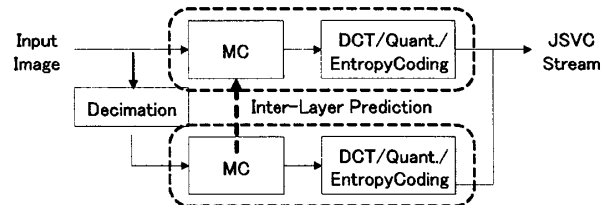


図 1: JSVC のブロック図

の縮小処理によりあらかじめ作成する。1 解像度を 1 レイヤとみなし、下位レイヤから順に各レイヤに H.264/AVC を適用する。そして、各レイヤで得られた符号化情報を多重化する。レイヤ間に発生する冗長性は、符号化済みの下位レイヤの符号化情報から上位レイヤの符号化情報を予測するレイヤ間予測により低減される。

### 2.1 JSVC の重み付き予測方法

P スライスおよび B スライスの参照フレームを 1 枚だけ用いる L0/L1 予測では、重み付き予測信号  $z$  を次のように算出する。

$$z = w \cdot y + d \quad (1)$$

ここで、 $y$  は動き予測参照信号である。 $w$  と  $d$  は WP パラメータであり、それぞれ重み係数とオフセット係数を示している。同様に、B スライスの参照フレームを 2 枚用いる双予測では、次のように算出する。

$$z = w_0 \cdot y_0 + w_1 \cdot y_1 + d \quad (2)$$

$y_0, y_1$  は動き予測参照信号、 $w_0, w_1$  は重み係数である。

P スライスおよび B スライス L0/L1 予測の重み付き予測は、Explicit モードしかとることができない。一方、B スライス双予測では、両モードを選択することができる。Explicit モードの場合には、 $w_0, w_1, d$  がシンタクスに埋め込まれる。Implicit モードの場合には、次のように符号化対象フレームと参照フレームとのフレーム間距離の比に応じて重み係数が決定される。

$$w_0 = 1 - w_1 \quad (3)$$

$$w_1 = t_b / t_d \quad (4)$$

$$d = 0 \quad (5)$$

フレーム間距離  $t_b, t_d$  のとり方は、図 2 を参照されたい。

2.2 H.264/AVC Implicit モード流用の問題点

Implicit モードは、間接的に WP パラメータを算出することで、余分な符号化情報を伝送しない方式である。したがって、算出される WP パラメータが輝度変化を十分に反映するものであれば、非常に効果的な手法となるが、それらが乖離すると予測性能は低下する。

前述した JSVC の重み付き予測 Implicit モードは、予測参照信号に対して図 2(b) に示すように重み係数のみによる線形予測を行う。したがって、仮にフレーム間の輝度変化が図 2(a) に示すように非線形であると、その分だけ原信号と予測信号は乖離する。非線形のフェード画像やフラッシュ画像が、これに該当する。この乖離を補正するために、双予測における重み付き推定ではオフセット係数  $d$  を設けている。しかし、JSVC の Implicit モードではこのオフセット係数がゼロにセットされるため、乖離量がそのまま予測残差として表れてしまう。

3 レイヤ間相関を利用した重み付き予測方法

3.1 本予測方法の狙い

現在の JSVC の Implicit モードは、H.264/AVC の機構をそのまま流用しているため、レイヤ内に閉じた時間的な輝度変化情報のみを用いて WP パラメータを算出している。しかし、JSVC では符号化済みの他レイヤの情報も参照可能である。そこで、本研究では、時間方向の輝度変化情報に加えて、直下レイヤの復号信号の情報を利用して WP パラメータを算出する。具体的には、直下レイヤの復号信号を利用してオフセット係数を算出する。重み係数は、従来の JSVC の Implicit モードと同様に参照フレームからのフレーム間距離の比で表現する。そして、図 2(c) に示すように、この直下レイヤの信号を利用して算出したオフセット係数により、線形予測からの乖離を補正し、予測残差を低減する。

3.2 WP パラメータの算出方法

$M \times N$  サイズの当該符号化対象ブロックの座標  $(m, n)$  における原信号を  $s^{mn}$  とおく。ブロック最左上の画素位置を原点と定めた座標系を想定する。 $M \times N$  は動き予測の算出単位であり、JSVC の場合、 $16 \times 16$ 、 $16 \times 8$ 、 $8 \times 16$ 、 $8 \times 8$ 、 $8 \times 4$ 、 $4 \times 8$ 、 $4 \times 4$  の 7 種類が考えられる。同様に双予測における 2 つの動き予測参照ブロックの座標  $(m, n)$  における復号信号を、 $y_0^{mn}$ 、 $y_1^{mn}$  とおく。この符号化対象ブロック一つにつき、一つの WP パラメータ  $w_0, w_1, d$  を割り当てるとする。

このとき、WP パラメータにより算出される重み付き予測信号と原信号の誤差  $e$  は、

$$e^{mn} = s^{mn} - (w_0 y_0^{mn} + w_1 y_1^{mn} + d) \quad (6)$$

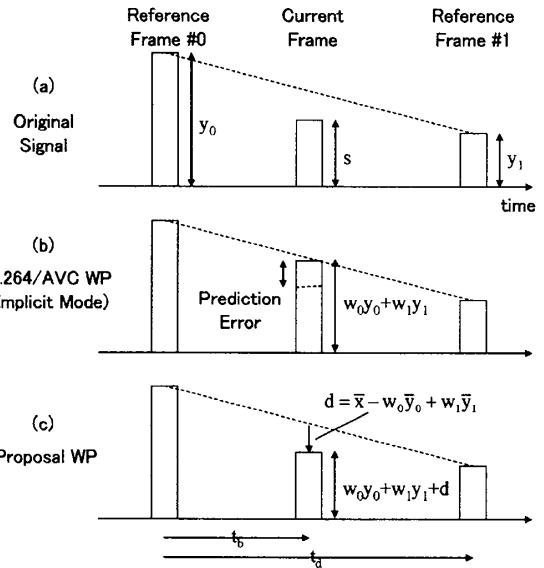


図 2: 従来手法と提案手法の重み付き予測信号の違い

として表現される。すると、当該ブロックにおける予測誤差  $e^{mn}$  の総エネルギー  $E$  は、

$$E = \sum_m \sum_n (s^{mn} - (w_0 y_0^{mn} + w_1 y_1^{mn} + d))^2 \quad (7)$$

として表現される。当該ブロックにおける最適な WP パラメータ  $w_0, w_1, d$  の値は、この総エネルギー  $E$  を最小化する値として考えることができる。ここで、重み係数  $w_0, w_1$  は、現在の Implicit モードの場合と同様に、式 (3)、式 (4) のようにフレーム間距離に応じて算出された値が一意に割り当てられると仮定する。すると、総エネルギー  $E$  を最小化するオフセット係数  $d$  は、 $\frac{\partial E}{\partial d} = 0$  を満たす値となる。これを解くと、次式のような

$$d = \bar{s} - w_0 \bar{y}_0 - w_1 \bar{y}_1 \quad (8)$$

ここで、 $\bar{s}$  は当該ブロックにおける原信号の平均値、 $\bar{y}_0, \bar{y}_1$  は動き予測参照ブロックにおける復号信号の平均値を指している。

このように、最適なオフセット係数は、当該ブロックの原信号の平均値と、動き予測参照ブロックの復号信号の平均値を用いた線形予測信号との差となる。しかし、復号器では原信号を参照できないため、別の情報で代替する必要がある。そこで、本手法では、原信号の平均値  $\bar{s}$  の代わりに、直下レイヤにおける当該ブロックと空間的同位置のブロックの復号信号の平均値  $\bar{x}$  を用いる。

$$d \approx \bar{x} - w_0 \bar{y}_0 - w_1 \bar{y}_1 \quad (9)$$

これは、レイヤ間の直流成分の保存性を利用している。解像度の縮小・拡大に用いるフィルタのフィルタ係数の

表 1: フェード画像に対する実験条件

Sequence	City, Foreman, Soccer
Num. of Frames	33 frames (white fade-in mixed)
Resolution	(BL) QCIF (EL) CIF
Frame Rate	30 fps
GOP	16 (Hierarchical B)
QP	21, 24, 27, 30
WP	(P-Slice) off (B-Slice) on

総和が1に正規化されていれば、各レイヤの空間的に対応するブロックの直流成分は保存される。したがって、直下レイヤの直流成分に含まれる量子化歪みが小さければ、原信号の平均値  $\bar{s}$  と直下レイヤの復号信号の平均値  $\bar{x}$  はほぼ等しくなる。

以上をまとめると、本提案 Implicit モードでは、重み係数の算出関数として、式 (3)、式 (4) を、オフセット係数の算出関数として、式 (9) を用いる。また、使用する Implicit モードを従来の JSVM か本提案手法か指定できるように、スライスヘッダ内に新たにフラグを設定した。これにより、スライス単位で切り替えを行うことができる。本稿では、スライス単位でコスト最小の方を適宜選択する2パス処理により符号化を実施した。

#### 4 実験

JSVC 参照ソフトウェア JSVM8.0 に提案手法を実装し、JSVM の Implicit モードとの比較を行った [3]。但し、両手法とも復号器ではマルチループデコードを行った。また、両手法とも動き推定を整数精度に限定して実施した。以上のような条件のもと、フェード画像およびフラッシュ画像に対して効果を検証した。

##### 4.1 フェード画像に対する検証結果

実験条件を表 1 に示す。3種類の JSVC 標準画像に33枚に渡る白のフェードインを付与した画像を使用した。白のフェード混合画像は次式のように作成した。

$$g(t) = (1 - f(t))a(t) + f(t)b(t) \quad (10)$$

ここで、 $g(t)$  はフェード混合画像、 $a(t)$  は白の単一色画像、 $b(t)$  は原画像、 $f(t)$  は混合係数の算出関数である。本実験では  $f(t)$  に、 $f(t) = t/T$  で変化する Linear Fade、および、 $f(t) = (t/T)^2$  で変化する Quadratic Fade の2種類を用意した。 $T$  は、フェード区間を指している。符号化対象はフェード部分の33枚とし、1フレームが1スライスに対応する。

実験結果を表 2 に示す。性能評価の指標として Bjontegaard Delta を使用した [4]。これは2つの RD 曲線

表 2: フェード画像に対する B ピクチャの符号化性能

Sequence	Fade	BD-bit [%]	BD-snr [dB]
City	Linear	-1.23	0.038
	Quadratic	-6.28	0.215
Foreman	Linear	-0.03	0.001
	Quadratic	-6.15	0.187
Soccer	Linear	-0.50	0.018
	Quadratic	-2.34	0.095

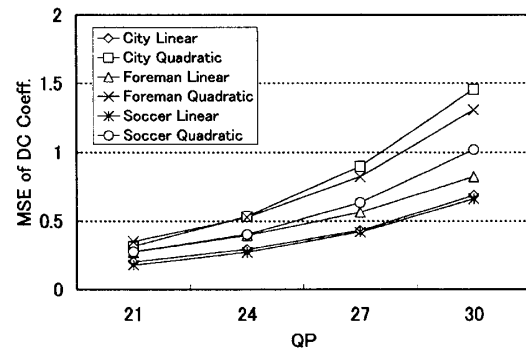


図 3: レイヤ間の直流成分の誤差

間の符号量と PSNR の差分値を計算する一手法である。表 2 では、B ピクチャにおける JSVM からの符号量増加率を BD-bit、PSNR 増加量を BD-snr としている。Linear fade の画像において、平均 0.59% の符号量削減、0.019dB の SN 改善、Quadratic fade の画像において、平均 4.92% の符号量削減、0.166dB の SN 改善を実現した。図 3 は、レイヤ間の対応するブロック間の直流成分の平均二乗誤差の QP に対する変化を示している。QP が増加すると誤差も増加するものの、その誤差は多くとも画素値にして1程度に収まる。つまり、式 (9) で仮定したレイヤ間の直流成分の保存性が妥当であることを示している。

図 4 に、Quadratic fade を付与した “City” における B ピクチャの BD-bit の時間的変化を、図 5 に、“City” のフェード混合画像の平均輝度値とその混合画像を QP24 で符号化した際の平均オフセット係数の時間的変化を示す。階層的 B ピクチャ構造をとるため、時間レベルが低い B ピクチャほど参照信号との輝度変化は線形から乖離し、本手法の効果が高くなる。図 4 の結果からも、時間レベルが低い B ピクチャにおいて、10%を超える大きな符号量削減が実現されていることが確認できる。また、図 5 より、そのような場合、オフセット係数の値も非ゼロの値が算出されており、オフセット係数が輝度変化を良好に補正できていることが裏付けられる。

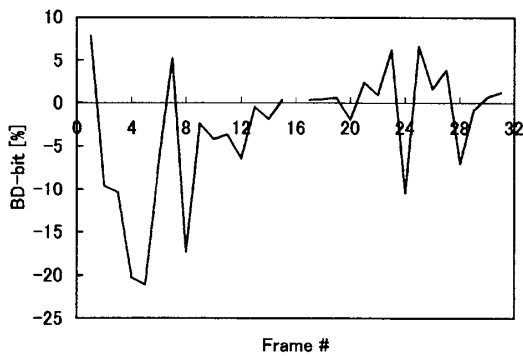


図 4: フェード画像における BD-bit の変化

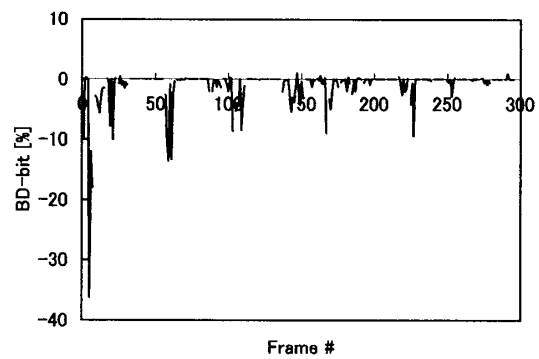


図 6: フラッシュ画像における BD-bit の変化

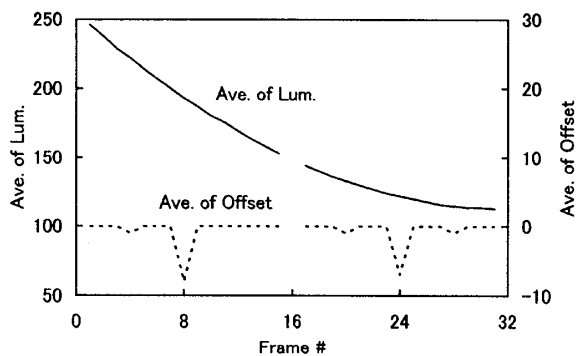


図 5: フェード画像における輝度値とオフセット係数

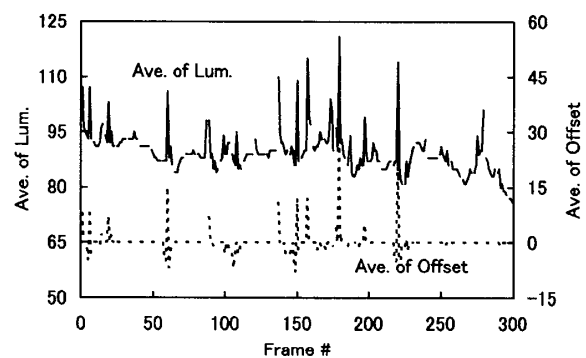


図 7: フラッシュ画像における輝度値とオフセット係数

#### 4.2 フラッシュ画像に対する検証結果

フラッシュ画像として JSVC 標準画像の Crew を使用し、300 枚の符号化を行った。GOP は 8 とした。その他の実験条件は、表 1 記載の条件で行った。

図 6 に、B ピクチャの BD-bit の時間的変化を、図 7 に、原画像の平均輝度値と QP24 で符号化した際の平均オフセット係数の時間的変化を示す。図 7 で、平均輝度値が急峻に変化しているフレームが、フラッシュを含むフレーム（以降、フラッシュフレーム）である。また、提案 Implicit モードを適用しないフレームのオフセット係数はゼロとしている。図 6 を見ると、フラッシュフレームとその近傍に対して、約 10% 程度（最大 36.2%）の符号量削減を実現していることが分かる。シーケンスの B ピクチャ全体の BD-bit は -1.14%、BD-snr は 0.043dB であった。輝度信号とオフセット係数を照合すると、BD-bit の推移と同様、フラッシュフレームとその近傍において、非ゼロのオフセット係数が算出されていることが分かる。フラッシュフレームに対しては、輝度の増加分だけ正のオフセット係数が割り当てられている。その近傍は、フラッシュフレームを予測参照しているフレームであり、輝度の高いフラッシュフレームの参照による線形予測からの乖離を、負のオフセット係数により補正し

ている。算出したオフセット係数が、フラッシュ画像に対しても良好に輝度変化を補正していることが分かる。

#### 5 おわりに

本稿では、マルチレイヤ構造をとる空間スケーラブル符号化における効率的な重み付き予測手法を提案した。直下レイヤの復号信号を参照することで、非線形の輝度変化を含むシーケンスに対して、予測性能を改善し、平均 4.92% の符号量削減を実現した。今後は、片方向予測におけるレイヤ相関を利用した WP パラメータ推定方法に取り組む予定である。

#### 参考文献

- [1] T. Wiegand, et.al.: "Joint Draft 8 of SVC Amendment," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-U201, 2006.
- [2] ITU-T: "Advanced video coding for generic audiovisual services," ITU-T Rec. H.264, 2003.
- [3] 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.
- [4] G. Bjontegaard: "Calculation of average PSNR differences between RD-curves," ITU-T SG16 Q.6 VCEG, VCEG-M33, 2001.