

ブロック歪みを考慮した時空間コントラスト感度特性に基づく

H.264/AVC 符号化器設計

H.264/AVC encoder design based on spatio-temporal contrast sensitivity considering block-artifact

坂東幸浩[†], 早瀬和也[†], 高村誠之[†], 上倉一人[†], 八島由幸[†]Yukihiro BANDO[†], Kazuya HAYASE[†], Seishi TAKAMURA[†], Kazuto KAMIKURA[†], and Yoshiyuki YASHIMA[†]

1 はじめに

高画質・高圧縮のH.264[1]符号化器を実現するためには、符号化器において適切な符号化パラメータを選択する必要がある。代表的なH.264符号化方式であるJM[2]およびJSVM[3]では、ラグランジェの未定乗数法に基づき、レート・歪みコストを最小化する予測モードを選択している。しかし、JMやJSVMで用いられている歪み量の尺度は二乗誤差であり、主観的な画質劣化を反映した歪み量とはいえない。例えば、二乗誤差は同一値であっても、高周波数成分の変化は低周波成分の変化に比べて、視覚的には検知されにくいことが知られている[4]。

そこで、本稿では、こうした視覚特性を考慮して、コントラスト感度特性に基づく歪み尺度をレート・歪みコストに導入し、効率的に符号量を削減する符号化パラメータ選択方法を検討する。本手法の基本的な考え方は、周波数成分に対する感度の違いをコントラスト感度関数によりモデル化し、歪み尺度を設定することである。ただし、その際、以下の2点を考慮する。

まず、ブロック境界における不連続性が主観画質に及ぼす影響を定量的に評価する。これは、動き補償によるフレーム間予測と直交変換を組み合わせたブロックベースの符号化方式では、ブロック歪みと呼ぶブロック境界における不連続性が特徴的な符号化歪みとして発生するためである。コントラスト感度関数に基づく歪み量への重み付けを各ブロックに対して独立に行う方法[5][6][7]では、ブロック歪みが考慮されず、得られた重み付き歪み量が主観画質を正しく反映できないケースが発生する。これは、重み付け処理をブロック内に閉じて行ったことが原因である。つまり、ブロック間の依存関係として、ブロック境界における不連続性についても評価の対象に加える必要がある。

さらに、空間周波数成分および動きの方向依存性を考慮した適応処理を導入する。これは、時空間領域における視覚系の検知機構が空間的エッジ方向および動きの方向に依存するためである。例えば、縦縞が移動する場合、動きとして認識できるのは、動きの水平成分のみとなる。つまり、

時空間周波数に対する視覚感度を評価する際は、空間周波数成分のエッジ方向および変移方向を考慮する必要がある。

2 時空間視覚感度に基づく歪み尺度の算出

ブロック内の歪み量に対して、コントラスト感度関数に基づく重み付けを行う。この重み付けは、DCT係数の量子化誤差に感度係数と呼ぶ重み係数を乗じることで実現する。コントラスト感度関数は、周波数成分毎の相対的感度を表したものであり、同関数を利用するためには、対象となる信号を周波数分析する必要がある。そこで、まず、2.1では、ブロック歪みを考慮した周波数分析について説明する。次に、2.2では、2.1の周波数分析の結果をコントラスト感度関数に適用する方法について詳述する。最後に、2.3では、2.2で得られるコントラスト感度を時空間周波数の視覚感度として拡張する方法について示す。この結果、時空間周波数の方向依存性を考慮した時空間感度を設定することが可能となる。

2.1 ブロック歪みを考慮した周波数分析

変換行列 $\Phi(N \times N$ 行列)の第 k 列ベクトル(N 次元ベクトル)を ϕ_k とすると、同行列に対する基底画像は、次式より得られる。なお、H.264の場合、 N として取りうる値は4または8のいずれかである。

$$f_{k,l}(x,y) = \phi_k[y]\phi_l[x]^t \quad (0 \leq x,y \leq N-1)$$

ここで、 ϕ_l^t は ϕ_l の転置ベクトルである。イントラないしインター予測により得られる予測誤差信号として、 $N \times N$ [画素]の領域における予測誤差信号 $R_m[x,y]$ ($N_{i_x} \leq x \leq N_{i_x} + N - 1, N_{i_y} \leq y \leq N_{i_y} + N - 1$)を $R_m^{(i_x,i_y)}$ と略記し、以降、基準ブロックと呼ぶ。さらに、対応する直交変換係数を $C_m^{(i_x,i_y)}[k,l]$ ($0 \leq k,l \leq N-1$)とすると、同予測誤差信号は次式のように表せる。

$$R_m^{(i_x,i_y)} = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} C_m^{(i_x,i_y)}[k,l] f_{k,l} \quad (1)$$

$M_x \times M_y$ 個の基準ブロック($R_m^{(i_x,i_y)}$)から構成される $M_x \times M_y \times N$ 画素を含む予測誤差信号 $R_m[x,y]$ ($N_{i_x0} \leq x \leq N_{i_x0} +$

[†]日本電信電話株式会社 NTT サイバースペース研究所

$M_x - 1, N_{i_0} \leq y \leq N(i_0 + M_y) - 1$) (分析対象ブロックと呼ぶ) に対して、ブロック歪を考慮した主観画質を考察する。分析対象ブロックと各基準ブロックの関係を表すためには、 $R_m^{(i_x, i_y)}$ に対して、ゼロ埋めにより、 $M_x N \times M_y N$ 画素の信号を得る。

$$\tilde{R}_m^{(i_x, i_y)}(x, y) = \begin{cases} R_m^{(i_x, i_y)}(x, y) & (N_{i_x} \leq x \leq N_{i_x} + N - 1, \\ & N_{i_y} \leq y \leq N_{i_y} + N - 1) \\ 0 & (\text{otherwise}) \end{cases}$$

同様に、各基底画像 $f_{k,l}(x, y)$ ($0 \leq x, y \leq N - 1$) に対して、ゼロ埋めにより、 $M_x N \times M_y N$ 画素の信号を得る。

$$\tilde{f}_{k,l}^{(i_x, i_y)}(x, y) = \begin{cases} f_{k,l}(x, y) & (N_{i_x} \leq x \leq N_{i_x} + N - 1, \\ & N_{i_y} \leq y \leq N_{i_y} + N - 1) \\ 0 & (\text{otherwise}) \end{cases}$$

ゼロ埋めの結果得られる $\tilde{f}_{k,l}^{(i_x, i_y)}(x, y)$ ($x = 0, \dots, M_x N - 1; y = 0, \dots, M_y N - 1$) を修正基底画像と呼ぶ。このとき、分析対象ブロックは、修正基底画像を用いて次式のように表すことができる。

$$\sum_{i_x=i_{x_0}}^{i_{x_0}+M_x-1} \sum_{i_y=i_{y_0}}^{i_{y_0}+M_y-1} \tilde{R}_m^{(i_x, i_y)} = \sum_{i_x=i_{x_0}}^{i_{x_0}+M_x-1} \sum_{i_y=i_{y_0}}^{i_{y_0}+M_y-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} C_m^{(i_x, i_y)} [k, l] \tilde{f}_{k,l}^{(i_x, i_y)}$$

両辺のフーリエ変換 ($\mathcal{F}[\cdot]$ で表記) は、フーリエ変換の線形性より、次式のように表すことができる。

$$\begin{aligned} \mathcal{F} \left[\sum_{i_x=i_{x_0}}^{i_{x_0}+M_x-1} \sum_{i_y=i_{y_0}}^{i_{y_0}+M_y-1} \tilde{R}_m^{(i_x, i_y)} \right] &= \sum_{i_x=i_{x_0}}^{i_{x_0}+M_x-1} \sum_{i_y=i_{y_0}}^{i_{y_0}+M_y-1} \mathcal{F} [\tilde{R}_m^{(i_x, i_y)}] \\ &= \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \left\{ \sum_{i_x=i_{x_0}}^{i_{x_0}+M_x-1} \sum_{i_y=i_{y_0}}^{i_{y_0}+M_y-1} C_m^{(i_x, i_y)} [k, l] \mathcal{F} [\tilde{f}_{k,l}^{(i_x, i_y)}] \right\} \end{aligned}$$

$\mathcal{F} [\tilde{f}_{k,l}^{(i_x, i_y)}]$ は、 $M_x N \times M_y N$ 次元の複素ベクトルであり、その第 (u_x, u_y) 要素は $F_{k,l}^{(i_x, i_y)}(u_x, u_y)$ として表すフーリエ係数である。 u_x, u_y を空間周波数インデックスと呼ぶ。フーリエ変換の対象を $f_{k,l}(x, y)$ ではなく、 $\tilde{f}_{k,l}^{(i_x, i_y)}(x, y)$ とすることで、ブロックの不連続性に起因する周波数成分も考慮した周波数分析を行うことが可能となる。

2.2 コントラスト感度関数の設定

2.1 で得られたフーリエ係数に対して、視覚感度に基づく重み付けを行う。この重み付けには、次式に示すコントラスト感度関数を用いる [4]。

$$g(\eta) = (0.31 + 0.69\eta) \exp(-0.29\eta) \quad (2)$$

ここで、 η は単位視野角内の明暗対の個数を表す空間周波数 (cycle per degree) である。また、 η は一次元の空間周波数である事に注意する。

このとき、 η と空間周波数インデックス u (u_x または u_y のいずれか) とは、以下の関係にある。

$$\eta(u, r) = \frac{\theta(r, H)u}{2MN} \quad (3)$$

ここで、 $u = u_x$ の場合、 $M = M_x$ であり、 $u = u_y$ の場合、 $M = M_y$ である。また、 $\theta(r, H)$ は縦幅 H の画像を視距離 rH において観測する場合の画素あたりの角度 (degree per pixel) であり、次式により与えられる。

$$\theta(r) = \frac{1}{H} \arctan\left(\frac{1}{r}\right) \frac{180}{\pi} \quad (4)$$

なお、画像の縦幅 H は、画素ピッチあたりの長さを α とし、対応する画素数を h とすると、 $H = \alpha h$ として表される。

式 (3) および式 (4) を式 (2) に代入することで、コントラスト感度関数 $g()$ は、パラメータ r の関数となっていることが確認できる。

2.3 感度係数の設定

空間周波数に対するコントラスト感度は、時間軸方向の変化量に依存する。例えば、時間軸方向の大きな変化 (高速なカメラパン・チルト、シーンチェンジ等) が発生したフレームでは、そのフレームの画質劣化に対する感度は低下する。このため、時間軸方向の大きな変化が発生したフレームに対しては、感度係数が小さな値になるよう制御することで、さらなる符号量削減が期待できる。コントラスト感度と視距離パラメータの関係に着目すると、2.2 の導出結果より、コントラスト感度は視距離パラメータ r の減少関数であることがわかる。そこで、本稿では、 r を時間方向の重みの制御パラメータとして用いることとし、次式のように変移量 d の大きさに応じて適応的に変化させるものとする。

$$r(d) = \begin{cases} r_1 & (\text{if } |d| \geq A) \\ r_2 & (\text{otherwise}) \end{cases} \quad (5)$$

ここで、 A は閾値であり、 $r_1 > r_2$ とする。以下では、式 (5)、式 (4) および式 (3) を式 (2) に代入し、 $g(\eta)$ を u, d の関数として表した次式の $\hat{g}(u, d)$ を時空間視覚感度関数と呼ぶ。

$$\hat{g}(u, d) = g(\eta(u, r(d))) \quad (6)$$

この時空間視覚感度関数を用いて、次式のようにフーリエ係数 $F_{k,l}^{(i_x, i_y)}(u_x, u_y)$ ($0 \leq u_x \leq M_x N - 1, 0 \leq u_y \leq M_y N - 1$) に対する重み付けを行う。

$$\tilde{F}_{k,l}^{(i_x, i_y)}(u_x, u_y, d_x, d_y) = F_{k,l}^{(i_x, i_y)}(u_x, u_y) \hat{g}(u_x, |d_x|) \hat{g}(u_y, |d_y|) \quad (7)$$

ここで、 (d_x, d_y) は、分析対象ブロックに対して推定された動きベクトルとする。なお、 $\hat{g}(u_x, |d_x|)$ および $\hat{g}(u_y, |d_y|)$ は、

各々、水平・垂直方向の重み関数となっている。式(7)は視覚系の検知機能がエッジに対して垂直な方向の動きに反応することに基づいたものである。動きの水平成分は水平エッジからなるテクスチャ領域では検知されず、同様に、動きの垂直成分は垂直エッジからなるテクスチャ領域では検知されない。離散フーリエ変換の空間周波数インデックス u_x は、水平方向の変化の度合いを表す。このため、空間周波数インデックス u_x の表す空間周波数に対応する波形は、垂直方向に移動しても、その動きは検出されず、動きが検出されるのは水平方向の移動に限定される。したがって、 u_x の表す空間周波数に対する時空間周波数の観点からの重みは、動きベクトルの水平方向成分 d_x と関連付ける必要があることから、 $\hat{g}(u_x, |d_x|)$ を用いている。 $\hat{g}(u_y, |d_y|)$ についても同様である。

分析対象ブロック内の信号 $R_m^{(i_x, i_y)}$ の (k, l) 基底の成分に対する感度係数を次式の電力比として定義する。

$$W_{k,l}(d_x, d_y) = \frac{\sum_{i_x=i_{x0}}^{i_{x0}+M_x-1} \sum_{i_y=i_{y0}}^{i_{y0}+M_y-1} C_m^{(i_x, i_y)} [k, l]^2 \tilde{P}_{k,l}^{(i_x, i_y)}(d_x, d_y)}{\sum_{i_x=i_{x0}}^{i_{x0}+M_x-1} \sum_{i_y=i_{y0}}^{i_{y0}+M_y-1} C_m^{(i_x, i_y)} [k, l]^2 P_{k,l}^{(i_x, i_y)}(u_x, u_y)}$$

ここで、 $\tilde{P}_{k,l}^{(i_x, i_y)}(d_x, d_y)$ 、 $P_{k,l}^{(i_x, i_y)}(u_x, u_y)$ は、以下の通りである。

$$\tilde{P}_{k,l}^{(i_x, i_y)}(d_x, d_y) = \sum_{u_x=0}^{M_x N-1} \sum_{u_y=0}^{M_y N-1} \tilde{F}_{k,l}^{(i_x, i_y)}(u_x, u_y, d_x, d_y)^2$$

$$P_{k,l}^{(i_x, i_y)}(u_x, u_y) = \sum_{u_x=0}^{M_x N-1} \sum_{u_y=0}^{M_y N-1} F_{k,l}^{(i_x, i_y)}(u_x, u_y)^2$$

上述の感度係数を用いて、第 (k, l) 成分の量子化誤差電力を $e_{k,l}$ として、本稿で考察する歪み量 \tilde{D} を定義する。

$$\tilde{D}(d_x, d_y) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} e_{k,l} W_{k,l}(d_x, d_y) \quad (8)$$

3 実験

本手法を参照ソフトウェア JSVM(version 8.0.1[3]) に実装し、無改造の JSVM との比較実験を行った。本手法の特徴は、新歪み尺度(式(8))に基づくラグランジュコストを最小化する符号化パラメータ選択にある。なお、JSVM は JSVC の参照ソフトウェアであり、基本階層は H.264 準拠の符号化ストリームを出力する。このため、単一階層の符号化を行う場合、H.264 復号器で復号可能な符号化ストリームを出力する。本章の実験結果は、本手法および無改造の JSVM を単一階層の符号化に適用したものである。

表 1: JSVM 符号化条件

設定項目	設定値
GOP structure	IPPP ...
QP of ISlice/PSlice	28
Hadamard transform	used
ME Search range	64
Total number of references	5
Entropy coding method	CABAC
Motion Estimation Scheme	Full Search
RD-optimized mode decision	used

表 2: JSVM および本手法のビットレート (括弧内は式(9)より求まる符号量削減率)

Sequences	JSVM [kbps]	Ours [kbps]
City	453	425 (-6.24 %)
Soccer	628	601 (-4.31 %)

3.1 実験条件

実験条件を表 1 に示す。符号化対象のシーケンスは、サイズ 352×288 [pixels]、フレームレート 30 [fps] である。符号化処理は、先頭の 120 フレームに対して実施した。式(5)におけるパラメータは $r_1 = 8$, $r_2 = 6$, $A = 5$ とした。また、基準ブロックのサイズを与えるパラメータは $N = 4$ とし、分析対象ブロックのサイズを与えるパラメータは $M_x = M_y = 4$ とした。

3.2 符号化効率

符号量の比較結果を表 2 に示す。いずれのシーケンス、QP 値においても提案手法によって、符号量の削減が図られていることが確認できる。なお、両手法の復号画像には、主観的な画質の差が認められない事を確認している。さらに、JSVM に対する提案手法の相対的な符号量削減率を評価するため、JSVM の符号量および提案手法の符号量を各々 R_{JSVM} , R_{Ours} として、次式で示した符号量削減率を表 2 の 3 列目の括弧内に示す。

$$\frac{R_{Ours} - R_{JSVM}}{R_{JSVM}} \times 100 [\%] \quad (9)$$

この結果、本手法は、JSVM に対して平均 5.3% の符号量低減を実現していることが確認できる。

表 3 に符号化モードの割合を示す。ここで、SKIP の列は、スキップモードの割合を示しており、INTER の列は、

スキップモードを除くインター予測の割合を示している。また、INTRAの列は、全てのイントラ予測モードの選択された割合を示している。同表が示すように、本手法のビットレートの削減は、発生符号量の少ないスキップモードを多く選択することで実現していることが分かる。

3.3 ブロック歪みの評価

さらに、ブロック歪みについての評価結果を示す。各フレーム ($X \times Y$ [画素]) の位置 (x, y) における復号画素値を $S(x, y)$ とし、水平・垂直方向の画素間差分値を各々

$$\delta_h(x, y) = S(x+1, y) - S(x, y)$$

$$\delta_v(x, y) = S(x, y+1) - S(x, y)$$

とする。このとき、次式に示すブロック境界における画素間差分値の平均値 Δ を用いて、隣接ブロック間の不連続性を評価する。

$$\Delta = \frac{\sum_{i_x=1}^{X/N-1} \sum_{i_y=0}^{Y-1} |\delta_h(Ni_x, i_y)|}{2(X/N-1)Y} + \frac{\sum_{i_y=1}^{Y/N-1} \sum_{i_x=0}^{X-1} |\delta_v(i_x, Ni_y)|}{2X(Y/N-1)}$$

表4にJSVMおよび本手法に対する Δ の全フレーム平均値を示す。なお、各行の3列目の値と2列目の値を各々、 Δ_{JSVM} , Δ_{Ours} とし、次式で定義するブロック歪み削減率を3列目の括弧内に示す。

$$\frac{\Delta_{Ours} - \Delta_{JSVM}}{\Delta_{JSVM}} \times 100 \text{ [%]} \quad (10)$$

3.4 考察

このように本手法は、JSVMに対して符号量を低減し、さらに、ブロック歪みも低減できていることが確認できる。この理由は以下のように考察される。本手法では、隣接ブロック間の依存関係を考慮した周波数分析を行っている。このため、隣接ブロック間の不連続性は、水平・垂直方向のエッジとして、コントラスト感度関数が大きな重みを与える周波数成分として分析される。つまり、大きな感度係数の値が設定される。この感度係数により重み付けられた歪尺度がコスト関数に用いられるため、ブロック間の不連続性をもたらす成分の歪みは回避され、結果として表4に示すように、ブロック歪みが低減したものと考察される。

4 おわりに

本報告では、符号化歪みの主観画質への影響を考慮した符号化パラメータ選択法について検討した。計算機シミュレーションの結果、主観画質を保持しつつ、符号量を平均5.3%低減できることを確認した。さらに、符号量の低減をはかり、かつ、ブロック歪みも低減可能であることを確認した。

表3: 符号化時の符号化モードの割合 [%]

(a) City			
	SKIP	INTER	INTRA
JSVM	38.41	60.10	1.44
Ours	47.50	51.08	1.42

(b) Soccer			
	SKIP	INTER	INTRA
JSVM	44.46	52.91	2.62
Ours	50.74	46.74	2.53

表4: JSVM および本手法のブロック間の不連続性 (括弧内は式(10)より求まるブロック歪み削減率)

Sequences	JSVM	Ours
City	1.81	1.79 (-0.69%)
Soccer	1.31	1.29 (-1.75%)

参考文献

- [1] "Advanced video coding for generic audiovisual services", ITU-T Rec. H.264 and ISO/IEC 14496-10 AVC, 2005.
- [2] K. P. Lim, G. Sullivan and T. Wiegand, "Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods", JVT-R95, Jan., 2006.
- [3] J. Reichel, H. Schwarz and M. Wien, "Joint Scalable Video Model JSVM-8". JVT-U202, Oct., 2006.
- [4] K. N. Ngan, K. S. Leong, and H. Singh. "Cosine transform coding incorporating human visual system model". *SPIE Fiber*, pp. 165-171, Sept., 1986.
- [5] A. B. Watson, "DCT quantization matrices visually optimized for individual images", *Proc. Hum. Vis., Visual Process., Digital Display IV*, pp. 202-216, 1993.
- [6] H. B. Yin and X. Z. Fang, "Adaptive quantization in perceptual MPEG video encoders", In *Proc. PCS2006*, Beijing, China, P3-14, Apr., 2006.
- [7] C. Jianwen, Z. Jianhua and H. Yun, "Macroblock-Level Adaptive Frequency Weighting for Perceptual Video Coding", *IEEE Trans. on Consumer Electronics*, vol. 53, no. 2, pp. 775-781, May, 2007.