社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－AVM－54 (7)
2006／9／28

# 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

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
〒239-0847 神奈川県横須賀市光の丘 1-1
E-mail : {bandou.yukihiro, takamura.seishi, kamikura.kazuto, yashima.yoshiyuki}@lab.ntt.co.jp

**あらまし** 高画質・高圧縮の H.264 コーデックを実現するためには、エンコーダにおいて適切な予測モードを選択する必要がある。例えば、代表的な H.264 エンコーダである参照ソフトウェア JM 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 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なっていない。例えば、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。また、動きの早いシーンにおける画質劣化は静止シーンにおける画質劣化よりも相対的に目立ちにくいことが知られている。そこで、本稿では、こうした視覚特性を考慮して、主観画質を反映した歪み量を用いることにより、効率的に符号量を削減する予測モード選択方法を検討する。

# A study on H.264/AVC mode decision based on human visual system

Yukihiro BANDOH†, Seishi TAKAMURA†, Kazuto KAMIKURA†, and Yoshiyuki YASHIMA†

†NTT Cyber Space Laboratories, NTT Corporation
1-1 Hikarino-oka, Yokosuka, Kanagawa 239-0847, JAPAN
E-mail : {bandou.yukihiro, takamura.seishi, kamikura.kazuto, yashima.yoshiyuki}@lab.ntt.co.jp

**Abstract:**
It is really important to decide proper prediction mode in H.264 encoder, since there much more modes than conventional method such as MPEG-2. Reference software (JM) of H.264 decides prediction mode based on the Lagrange's method of undetermined multipliers where squared error is used as the criterion of distortion. However, squared error dose not always correspond to the distortion from the viewpoint of subjective quality. In this paper, we investigate the mode decision method based on human visual system in order to improve H.264 encoder.

# 1 はじめに

映像符号化の国際標準規格 H.264/AVC[1] が、近年、大きな注目を集めている。その用途はワンセグ放送、3GPP における携帯向けマルチメディアサービス、各種衛星放送、欧州におけるデジタル放送 (The Digital Video Broadcast (DVB))、次世代 DVD フォーマットのように、携帯端末における映像を対象とした低ビットレートから、HDTV クラス以上の映像を対象とした高ビットレートに至るまで広範囲にわたる。こうした期待の高まりは、ひとえに、H.264/AVC のポテンシャルの高さがもたらしたものである。H.264 に高い符号化性能を付与する要素技術としては、予測符号化、エントロピ符号化、デブロッキングフィルタの 3 つをあげることができる。とりわけ、予測符号化に関連するモード選択は、従来の標準化方式 (MPEG-2, MPEG-4, ect) と比べても、符号化器における設計の自由度が高い。これは、イントラ予測、可変形状動き補償および複数フレーム参照の導入により、H.264 では、予測モードの種類が増加しているためである。

このため、高画質・高圧縮の H.264 符号化器を実現するためには、符号化器において適切な予測モードを選択する必要がある。こうした符号化器での動作は標準化の範囲外であり、設計の自由度が残されている。例えば、代表的な H.264 符号化器である参照ソフトウェア JM[2] 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。このとき、JM で用いられている歪み量の尺度は二乗誤差である。

しかし、二乗誤差は必ずしも、主観的な画質劣化を反映した歪み量ではない。例えば、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 [3] [4] [5] [6] 。また、動きの早いシーンにおける画質劣化は静止シーンにおける画質劣化よりも相対的に目立ちにくいことが知られている（時間マスキング効果)。このため、こうした視覚特性を利用していない JM には、符号量の効率的な削減に関して、改良の余地が残る。

そこで、本稿では、時空間周波数の視覚感度およびマスキング効果を考慮して、主観画質を反映した歪み量を用いることにより、効率的に符号量を削減する予測モード選択方法を検討する。

# 2 JM のモード選択におけるコスト関数

参照ソフトウェア JM では、以下のレート・歪みコストを最小化する予測モードを選択している。

$$J(S,\hat{S}_{m,q},m,q,\lambda)=D(S,\hat{S}_{m,q})+\lambda R(S,\hat{S}_{m,q},m,q)$$

ここで、$S$ は原信号、$q$ は量子化パラメータ、$m$ は予測モードを表す番号であり、$\hat{S}_{m,q}$ は $S$ に対してモード $m$ を用いて予測し、$q$ を用いて量子化した場合の復号信号である。また、$\lambda$ はモード選択に用いるラグランジェの未定乗数である。さらに、$D(S,\hat{S}_{m,q})$ は次式に示す二乗誤差和である。

$$\begin{aligned} D(S,\hat{S}_{m,q}) &= \sum_{x=0}^{15}\sum_{y=0}^{15}|S^{Y}[x,y]-\hat{S}^{Y}_{m,q}[x,y]|^2 \\ &+ \sum_{x=0}^{7}\sum_{y=0}^{7}|S^{U}[x,y]-\hat{S}^{U}_{m,q}[x,y]|^2 \\ &+ \sum_{x=0}^{7}\sum_{y=0}^{7}|S^{V}[x,y]-\hat{S}^{V}_{m,q}[x,y]|^2 \end{aligned}$$

ここで、$S^{Y}$, $S^{U}$, $S^{V}$ は原信号の Y,U,V 成分であり、$\hat{S}^{Y}_{m,q}$, $\hat{S}^{U}_{m,q}$, $\hat{S}^{V}_{m,q}$ は復号信号の Y,U,V 成分である。

# 3 時空間視覚感度に基づく歪み量の重み付け

空間周波数成分毎に主観画質への影響に差があるのは、空間周波数成分に対する視覚感度に差があることに由来する。そこで、符号化に用いる直交変換係数に対して、周波数分析を行い、周波数成分毎の視覚感度に応じて歪み量に重み付けを行うことで、主観画質を反映した歪み量を定義する。さらに、時間マスキングを考慮して、上述の重み付けされた歪み量に対して、動き量に応じて重み付けを行う。こうした時空間的な性質に応じて重み付けされた歪み量をレート・歪みコスト内で用いる。この歪み量の変更は、主観画質を最大限保持しつつ、符号量を低減するという符号化本来の目的に対して、より適切な解を与えることにつながる。

## 3.1 量子化誤差信号の重み付け

量子化誤差信号に対する視覚感度に基づく重み付けについて、以下、説明する。本手法では、次式の

レート・歪みコストを用いる。

$$\begin{aligned} & J(S,\hat{S}_{m,q},m,q,\lambda) \\ = \ & \tilde{D}(C_n,\hat{C}_q)+\lambda R(S,\hat{S}_{m,q},m,q) \end{aligned} \quad (1)$$

このレート・歪みコストの計算に用いる歪み量として、以下の重み付け歪み量を用いる。

$$\begin{aligned} & \tilde{D}(C_n,\hat{C}_q) \\ = & \sum_{i=0}^{\lfloor 16/N\rfloor\cdot\lfloor 16/N\rfloor-1}\sum_{k=0}^{N}\sum_{l=0}^{N}|C_n^{Y(i)}[k,l]-\hat{C}_q^{Y(i)}[k,l]|^2\cdot W_{k,l}^{Y} \\ + & \sum_{i=0}^{\lfloor 8/N\rfloor\cdot\lfloor 8/N\rfloor-1}\sum_{k=0}^{N}\sum_{l=0}^{N}|C_n^{U(i)}[k,l]-\hat{C}_q^{U(i)}[k,l]|^2\cdot W_{k,l}^{U} \\ + & \sum_{i=0}^{\lfloor 8/N\rfloor\cdot\lfloor 8/N\rfloor-1}\sum_{k=0}^{N}\sum_{l=0}^{N}|C_n^{V(i)}[k,l]-\hat{C}_q^{V(i)}[k,l]|^2\cdot W_{k,l}^{V} \end{aligned}$$

ここで、$C_n^{\gamma(i)}[k,l]$ $(\gamma=Y,U,V)$ はマクロブロック（Y成分の場合、16×16[画素], U,V成分の場合、8×8[画素]）内のサブブロック（$N\times N$[画素]）のうち、ラスター走査において $i$ 番目に走査されるサブブロックである。なお、$N$ は直交変換のサイズを表す変数であり、H.264の場合、$N$ の取りうる値は4または8のいずれかである。また、$C_q^{\gamma(i)}[k,l]$ $(\gamma=Y,U,V)$ はマクロブロック内の復号変換係数のうち、ラスター走査において $i$ 番目に走査されるサブブロックである。さらに、$W_{k,l}^{\gamma}$ $(\gamma=Y,U,V)$ は1以下に設定される重み係数であり、以下では、**感度係数**と呼ぶ。感度係数の算出については、次項にて詳述する。上式において、$W_{k,l}^{\gamma}$ を小さな値に設定することは、量子化歪み $D(C_n,\hat{C}_q)$ を小さく見積もることに相当する。なお、直交変換の正規性より、$W_{k,l}^{\gamma}=1$ $(\forall k,l;\gamma=Y,U,V)$ とすれば、上述の重み付け歪み量は二乗誤差和と等価となる。

## 3.2 感度係数の算出

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 $\phi_k$ とすると、同行列に対する基底画像は、次式より得られる。

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

ここで、$\phi_l{}^t$ は $\phi_l$ の転置ベクトルである。各基底画像 $f_{k,l}(x,y)$ $(0\le x,y\le N-1)$ に対して、次式に示す離散フーリエ変換を施し、フーリエ係数を得る。なお、次式において、$j$ は虚数単位である。

$$F_{k,l}(u,v)=\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f_{k,l}(x,y)\exp(-2\pi(\frac{xu}{N}+\frac{yv}{N})j) \quad (2)$$

得られたフーリエ係数 $F_{k,l}(u,v)$ $(0\le u\le N-1, 0\le v\le N-1)$ に対して、次式の通り、重み付けを行う。

$$\tilde{F}_{k,l}(u,v)=F_{k,l}(u,v)\hat{g}(\eta) \quad (3)$$

以下、$\tilde{F}_k(u,v)$ について、説明する。$\hat{g}(\eta)$ は $g(\eta)$ を用いた次式で表される関数である。

$$\hat{g}(\eta)=\begin{cases} 1 & (0\le\eta\le\eta_0) \\ \alpha g(\eta) & (\eta_0<\eta) \end{cases} \quad (4)$$

ここで、$\alpha$ は重みを制御するパラメータである。また、$g(\eta)$ は視覚感度関数として知られる関数であり、次式のような関数形で表される。

$$g(\eta)=(a+b\eta)\exp(-(c\eta)^d) \quad (5)$$

ここで、$a,b,c,d$ は視覚感度関数の関数形を定めるパラメータ（以後、**モデルパラメータ**と呼ぶ）であり、例えば、次のような値をとる。次の値は、上から順に、文献 [3] [4] [5] [6] による。

$$(a,b,c,d)=(0.4992,0.2964,-0.114,1.1) \quad (6)$$

$$(a,b,c,d)=(0.2,0.45,-0.18,1) \quad (7)$$

$$(a,b,c,d)=(0.31,0.69,-0.29,1) \quad (8)$$

$$(a,b,c,d)=(0.246,0.615,-0.25,1) \quad (9)$$

また、式 (5) において、$\eta$ は以下の値とする。

$$\eta=\frac{\theta(r,H)\sqrt{u^2+v^2}}{2N} \quad (10)$$

$\eta_0$ は以下のように、$g(\eta)$ が最大値をとる引数である。

$$\eta_0=\arg\max_{\eta}g(\eta) \quad (11)$$

$\theta(r,H)$ は縦幅 $H$ の画像を視距離 $rH$ において観測する場合の一画素あたりの角度であり、次式により与えられる。

$$\theta(r,H)=\frac{\arctan(\frac{1}{r})\frac{180}{\pi}}{H} \ [\text{degrees/pixel}] \quad (12)$$

以後、$r$ を**視距離パラメータ**と呼ぶ。

以上の議論より、$\tilde{F}_k(u,v)$ は視距離パラメータ $r$ の関数であることが分かる。そこで、以下では、$\tilde{F}_k(u,v)$

の代わりに $\tilde{F}_k(u,v,r)$ として、視距離パラメータ $r$ の関数であることを陽に示す表記法を用いる。

基底画像 $f_{k,l}(x,y)$ $(0 \le x,y \le N-1)$ に対する感度係数を次式の電力比として定義する。

$$W^Y_{k,l}(r) = \frac{\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\tilde{F}_{k,l}(u,v,r)}{\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}F_{k,l}(u,v)} \qquad (13)$$

感度係数は符号化対象画像とは独立に求めることが可能である。このため、符号化前に予め、感度係数を求め、ルックアップテーブルに格納すれば、符号化時の感度係数算出のための演算は省略することができる。なお、$W^U_{k,l}(r)$, $W^V_{k,l}(r)$ についても同様に求めることができる。

### 3.3 時間マスキング効果の影響を考慮した修正

あるフレームにおいて、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)が発生した場合、そのフレームの画質に対する感度は極端に低下する。これは、時間マスキング効果として知られる視覚特性である。そこで、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。

ここでは、$W^\gamma_{k,l}(r)$ と $r$ の関係に着目する。$W^\gamma_{k,l}(r)$ $(\gamma = Y,U,V)$ は次式を満たす。

$$W^\gamma_{k,l}(r) \le 1$$

これは、

$$g(\eta) \le 1$$

となることから、

$$\tilde{F}_k(u,v,r) \le F_{k,l}(u,v)$$

となるためである。また、$\tilde{F}_k(u,v,r)$ が $r$ の減少関数であることから、$W^\gamma_{k,l}(r)$ は視距離パラメータ $r$ に対する減少関数であることが分かる。これは、視距離と共に視覚感度が鈍化するという視覚特性に対応するものである。

そこで、本手法では、この感度係数の制御に視距離パラメータを用いる。この制御は、動きベクトルの大きさに応じてマクロブロック毎に行うものとする。このとき用いる動きベクトル $v = (v_x, v_y)$ は、動き推定のブロックサイズを $16 \times 16$ として求めるものとする。求めた動きベクトル $v$ に応じて、視距離パラメータを適応的に変化させる。

$$r = \begin{cases} r_1 & (|v_x| + |v_y| \ge A \text{ の場合}) \\ r_2 & (\text{otherwise}) \end{cases} \qquad (14)$$

ここで、$r_1 > r_2$ とする。

表 1: 符号化パラメータ (JM10.1 を使用)

| 設定項目 | 設定値 |
|---|---|
| GOP | IPPP |
| QP of ISlice | 22, 28, 34 |
| QP of PSlice | 22, 28, 34 |
| Hadamard transform | used |
| ME Search range | 64 |
| Total number of references | 5 |
| References for P slices | 5 |
| Entropy coding method | CABAC |
| Motion Estimation Scheme | Full Search |
| Search range restrictions | none |
| RD-optimized mode decision | used |
| Residue Color Transform | not used |

## 4 実験

本手法を参照ソフトウェア JM(version 10.1[2]) に実装し、デフォルトの JM との比較実験を行った。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 [pixles] の "foreman", "soccer", "bus" および、サイズ 720 × 480 [pixles] の "whale show","soccer", である。また、いずれのシーケンスもフレームレート 30[fps]、カラーフォーマット 4:2:0 である。なお、"whale show","soccer" のオリジナルは ITE の標準映像であり、このオリジナルのインターレース映像 (1080 lines) から top field のみを抜出し、水平方向を 1/2 に縮小した後、フレーム中央部の 720 × 480 [pixels] をクロッピングしたものである。視距離パラメータは、画面の高さを $H$ として、$r_1 = 5H$ (動ブロックの場合), $r_2 = 2H$ (非動ブロックの場合) とした。

符号量の比較結果を表 2 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

表 2: ビットレートの比較

(a) QP=22

| Sequences | JM [Kbps] | 提案法 [Kbps] |
|---|---|---|
| foreman | 948 | 872 |
| soccer1 | 1269 | 1209 |
| bus | 2554 | 2427 |
| whale show | 14114 | 12740 |
| soccer2 | 9725 | 8628 |

(b) QP=28

| Sequences | JM [Kbps] | 提案法 [Kbps] |
|---|---|---|
| foreman | 362 | 335 |
| soccer1 | 588 | 563 |
| bus | 1220 | 1163 |
| whale show | 7186 | 6422 |
| soccer2 | 3339 | 3102 |

(c) QP=34

| Sequences | JM [Kbps] | 提案法 [Kbps] |
|---|---|---|
| foreman | 151 | 146 |
| soccer1 | 266 | 258 |
| bus | 546 | 523 |
| whale show | 3294 | 2895 |
| soccer2 | 1226 | 1117 |

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、符号量の削減量に加えて、JM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、JMの符号量および提案手法の符号量を各々$R_{JM}$、$R_{Ours}$として、次式で示した符号量削減率を図1に示す。

$$100\frac{R_{JM}-R_{Ours}}{R_{JM}} \quad [\%]$$

この結果、本手法は、JMに対して平均7.7%(QP=22), 6.8%(QP=28), 6.3%(QP=34)の符号量低減を実現していることが確認できる。

## 5 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モード選択法について検討した。符号化実験の結果、主観画質を保持しつつ、符号量を平均6.3〜7.7%、低減できることを確認した。今後は、動き量に応じた視距離パラメータの設定の自動化について検討を加える予定である。

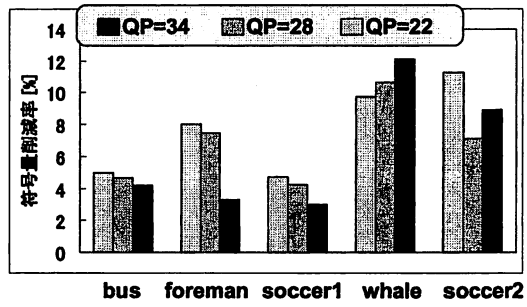


図 1: 符号量の削減率

## 参考文献


[1] Joint Video Team. *Draft ITU-T Recommendation and Final draft international standard of joint video specification.* ITU-T Rec.H.264 and ISO/IEC 14496-10 AVC, 2003.

[2] http://iphome.hhi.de/suehring/tml/download/jm10.1.zip .

[3] J.L.Mannos and D.J.Sakrison. The effect of a visual fidelity criterion on the encoding of images. *IEEE Trans. Infomation Theory*, Vol. IT-20, pp. 523–536, July 1974.

[4] 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.

[5] K.N.Ngan, K.S.Leong, and H.Singh. Cosine transform coding incorporating human visual system model. *SPIE Fiber*, pp. 165–171, Sept. 1986.

[6] B.Chitprasert and K.R.Rao. Human visual weighted progressive image transmission. *IEEE Trans. Commun.*, pp. 1040–1044, July 1990.