歪み最小化型ビット配分に基づくデジタル放送用 MPEG-2 符号化制御の最適化

Optimization of MPEG-2 Encoding Control Applicable for Digital Broadcasting with Full Use of Distortion Minimization Oriented Bit Assignment Technology

正会員 内藤 整[†], 正会員 小池 淳[†], 正会員 松本 修一[†], 正会員 羽鳥 好律^{††}

Sei Naito[†], Atsushi Koike[†], Shuichi Matsumoto[†] and Yoshinori Hatori^{††}

Abstract Distortion minimization oriented bit assignment technology is representative of approaches that have been studied to improve MPEG-2 rate control performance. This technology conducts an inclusive bit assignment for each picture while maintaining picture distortion at a minimum level. Its effectiveness has been tested and confirmed. In this paper, MPEG-2 HDTV encoding was assumed, and the distortion minimization oriented bit assignment technology was introduced as a basic scheme. We proposed additional improvement approaches with full use of the technology for improving macroblock quantization control and adaptive picture type selection. Coding experiments confirmed that an additional coding gain could be achieved with the proposed scheme, especially for low bit-rate encoding.

キーワード: 歪み最小化型ビット配分, 適応符号化制御, マクロブロック単位量子化制御, ピクチャタイプ

1. ま え が き

デジタル放送や DVD など, MPEG-2 をベースとする民生用の映像アプリケーションは, HDTV 化の流れにある。これらに代表されるような限られた伝送帯域・記録容量の条件下で, 従来型の MPEG-2 符号化制御により HDTV 品質を維持することは極めて困難である。これに対し, 低ビットレートの HDTV に限らず, MPEG-2 の枠組の中で, さらなる高効率化を実現する要素技術の検討例は既に多く見られ, 以下のように整理できる。

(1) ピクチャ構造

近傍の符号化結果に基づき, ピクチャ構造の適応選択を行う¹⁾。

(2) ビット配分

ピクチャごとのビット配分をレート-歪みの関係を考慮し, 歪み最小化の規範で行う²⁾³⁾。

(3) ピクチャタイプ

B ピクチャの挿入頻度を適応的に更新する。非リアルタイム系を想定したアプローチ⁴⁾とリアルタイム系を想定したアプローチ^{5)~7)}に大別される。

(4) マクロブロック符号化モード

マクロブロックごとの符号化方法を発生ビット数最小化の規範で適応的に選択する⁷⁾⁸⁾。

(5) 動きベクトル検出

動きベクトルの選択を単なる誤差最小化規範ではなく, コスト関数などの導入により符号化ビット数を考慮して行う^{9)~11)}。

筆者らは, 先にビット配分の最適化アプローチとして, 歪み最小化型のビット配分方式を提案した¹²⁾。導入による効果として, 符号化効率の向上に加え, 画面内のマクロブロックごとの優先度付けを, 量子化制御に忠実に反映する枠組を提供できるという意味で, 主観画質の改善に極めて有効であることを確認した。

以上の背景から, 本論文ではデジタル放送での応用を想定し, MPEG-2 による低ビットレート HDTV 符号化を対象とした高画質化アプローチとして, 歪み最小化型ビット配分の導入を前提に, 同手法との併用により性能改善が見込まれる適応符号化制御として, マクロブロック単位量子化制御の高精度化, ならびにピクチャタイプの最適割当に着目し, 方式検討を行った。

2002 年 2 月, 映像情報メディア学会放送方式研究会で発表

2003 年 2 月 7 日受付, 2003 年 6 月 5 日再受付, 2003 年 7 月 11 日採録

[†]株式会社 KDDI 研究所

(〒 356-8502 上福岡市大原 2-1-15, TEL 049-278-7412)

^{††}東京工業大学 大学院 総合理工学研究科

(〒 226-8502 横浜市緑区長津田町 4259, TEL 045-924-5528)

[†]KDDI R&D Laboratories Inc.

(2-1-15, Ohara, Kamifukuoka-shi, Saitama 356-8502, Japan)

^{††}Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

(4259, Nagatsudacho, Midori-ku, Yokohama, Kanagawa 226-8502, Japan)

本論文では、2章で、歪み最小化型ビット配分の導入を行う。3章では、歪み最小化型ビット配分の高度利用技術として提案する、マクロブロック単位量子化制御の高精度化、およびピクチャタイプの最適割当方式の詳細を述べる。4章では、符号化シミュレーション実験により提案方式の性能評価を行う。最後に5章で結論を述べる。

2. 前提条件 –歪み最小化型ビット配分の導入–

ビット配分の最適化アプローチとして、筆者らが先に提案した歪み最小化型ビット配分方式を以下に導入する。まずレート制御の導入目的は以下に集約される。

(A) コーデック処理の連続性保持

発生ビット数の長時間平均を一定量に抑え、デコーダ処理において破綻 (VBV バッファのオーバーフロー/アンダーフロー) を生じさせない。

(B) 一定符号化レート下での画質向上

(A) が満足される範囲内で、発生ビット数の割振りを時間方向、空間方向で最適化する。

このうち、歪み最小型ビット配分方式は (B) を主目的とするアプローチである。前提として、ビット配分は DCT 係数部分とサイド情報部分に分けて行うこととする。ここでピクチャタイプごとに消費するサイド情報のビット数については、シーン内での時間的変動が少ないことを確認している。よってシーン先頭を除き、当該ピクチャに対するサイド情報部分の割当は、近傍の同一ピクチャタイプの符号化結果から高精度に行うことが可能であるので、ビット配分の検討対象はピクチャに費やす全符号化ビットのうち DCT 係数部分に限定することとする²⁾³⁾。ただし、サイド情報とは符号化ビットのうち、量子化制御の対象にならない部分を指し、具体的には各種ヘッダ、動きベクトルに代表されるマクロブロック情報、Intra ブロックの DC 係数がこれに含まれる。

2.1 ピクチャレイヤのビット配分

まず、ピクチャタイプごとに R-D 関数 (レート–歪みの関係) を定式化し、式 (1) から式 (5) により、当該ピクチャの目標量子化誤差電力の算出を行う。

$$\begin{cases} D_I = f_I(R_I, \sigma_{xI}^2) \\ D_P = f_P(R_P, \sigma_{xP}^2) \\ D_B = f_B(R_B, \sigma_{xB}^2) \end{cases} \quad (1)$$

$$f(R, \sigma_x^2) = \gamma_x^2 \sigma_x^2 2^{-2mR}, 0 \leq \gamma_x^2 \leq 1 \quad (2)$$

$$D_I = D_P = \frac{1}{w_B} \times D_B \quad (3)$$

$$R_{SOP} = N_I \times R_I + N_P \times R_P + N_B \times R_B \quad (4)$$

$$m(R) = \frac{a}{R} + b \quad (a, b \text{ は実数}) \quad (5)$$

D_I, D_P, D_B	: 量子化誤差電力
R_I, R_P, R_B	: 目標符号化ビット数
$\sigma_{xI}^2, \sigma_{xP}^2, \sigma_{xB}^2$	: 量子化器入力信号の分散
f_I, f_P, f_B	: R-D 関数
γ_x^2	: 量子化器入力信号のスペクトル分布に応じて決まる定数
w_B	: B ピクチャの量子化誤差に対する重み係数
R_{SOP}	: SOP 内配分可能ビット数
N_I, N_P, N_B	: SOP 内残ピクチャ枚数

式 (1) 中、 (D_I, D_P, D_B) ならびに $(\sigma_{xI}^2, \sigma_{xP}^2, \sigma_{xB}^2)$ の算出対象は、ピクチャ内で量子化対象となるすべての輝度信号ブロックとし、 (D_I, D_P, D_B) は 1 画素あたりの値に換算して算出し、 $(\sigma_{xI}^2, \sigma_{xP}^2, \sigma_{xB}^2)$ は DCT ブロックごとに分散を求めた上で、ピクチャ平均により算出する。

式 (1) において関数 (f_I, f_P, f_B) は、ピクチャタイプごとの R-D 関数であり式 (2) により表現される¹³⁾。式 (2) において m は R-D 曲線の勾配を示すパラメータであり、素材別、ピクチャタイプ別に式 (5) により近似可能である¹²⁾。同関数を決定する実数 a, b の更新はピクチャの符号化を完了する都度、 D, R, σ_x^2 の実測値をもとに m 値を算出し、当該シーン内で既取得の m 値に対する誤差 2 乗和最小化の規範で行うこととした。式 (3) において $w_B (>1.0)$ はトータルの符号化効率を向上させる目的から、予測参照されない B ピクチャに許容する量子化誤差を I ピクチャ、および P ピクチャよりも高めに維持する重み係数である。式 (4) では、レート制御の区切りを表現するため、SOP という用語を用いた。MPEG-2 で定める GOP (Group of Pictures) が、単に GOP ヘッダおよび I ピクチャより始まる一連のピクチャ群を指すものであるのに対し、SOP は歪み最小化型ビット配分方式に特有のレート制御の区切りを示すという点で GOP とは区別される。参考までに Test Model 5¹⁴⁾ のレート制御では、SOP は GOP に一致する。SOP の割当ビット数 R_{SOP} の算出方法は、再生画質を安定的に維持する目的から、以下を考慮して定めることとした。

・符号化済ピクチャにおける目標ビット数に対する発生ビット数の過不足は、SOP 内残ピクチャの符号化により吸収するよう制御が働くため、SOP 内残ピクチャ枚数を余裕度として捉えた場合、残ピクチャ枚数はピクチャ位置に依らず一定であることが望ましい。

・1で述べたとおり、ビット配分による高画質化手法導入の前提となる (A) の要件を安定的に満たす上で、デコーダの VBV バッファ占有量を考慮した、ピクチャ単位の目標符号化ビット数の算出がデコーダ処理の破綻防止に極めて有効である。

具体的には、SOP 内残ピクチャ数が常に一定となるよう、SOP の割当領域をオーバーラップさせることとし

た¹⁵⁾。当該方式における SOP の扱いを図 1 に示す。ここで、ある符号化時刻 t における配分可能ビット数 R_{SOP} の算出は、式 (6) に従い行うこととする。 N_{SOP} は SOP 内ピクチャ数、つまり SOP 長を指し、GOP 長の整数倍とする。図 1 では SOP 長が GOP 長に等しい場合の例を示している。 i はシーケンス先頭ピクチャを 0 としたときのピクチャ位置を指し、 i 番目のピクチャにつき $R(i)$ 、 $S(i)$ はそれぞれレート制御上の目標ビット数、実際の発生ビット数を指す。同式は、当該ピクチャの符号化開始時における VBV バッファ占有量について実測値 (第 2 項) の理想値 v_T に対する過不足を、当該 SOP にて補償するという前提のもと R_{SOP} を算出するものである。また、VBV バッファ占有量の算出に関し、 $R(i)$ の部分には符号化レート相当の固定値を代入するのが通例である。 $R(i)$ 部分に固定値を与えた場合、シーケンス先頭では I ピクチャの存在からピクチャごとの発生ビット数が大きく異なり、結果として算出される占有量が不安定となる点を考慮し、同式では $R(i)$ 部分に該当するピクチャのレート制御時に算出した目標ビット数を与えることとした。VBV バッファ占有量の理想値としては、オーバフロー、アンダーフロー防止の観点から VBV バッファサイズの $1/2$ を適用することとした。

$$R_{SOP} = N_{SOP} \times \frac{\text{bit_rate}}{\text{frame_rate}} + \{v_0 + \sum_{i=0}^{t-1} (R(i) - S(i))\} - v_T \quad (6)$$

- t : シーケンス内符号化済ピクチャ枚数
 v_0 : シーケンス先頭でピクチャデータの引抜きが開始される VBV バッファ占有量
 v_T : 理想とする VBV バッファ占有量

2.2 マクロブロックレイヤのビット配分

マクロブロック単位ビット配分に関し、Test Model 5 に見られる仮想バッファをベースとする方式が代表的である。同手法が抱える問題点として以下が指摘されており、マクロブロックごとの符号化特性・視覚特性に応じた、柔軟なビット配分を実現する上で理想的な手法とはいえない。

図 1 SOP の割当
Assignment of SOP.

・マクロブロックごとの発生ビット数を平滑化する作用が働くため、細かい制御が実現できない。

・仮想バッファ占有量 d のピクチャ内の変動により、意図に反して再生画質に関する不均一性が画面内に生じる可能性がある¹⁶⁾。

以上の問題点を踏まえ、マクロブロックごとの優先度付けを忠実、かつ安定的に反映する目的から、マクロブロック単位ビット配分の改良手法を以下に導入する。

前提として、2.1 で導入した手法に準じ、当該ピクチャの符号化開始時には量子化誤差電力の目標値 (D_I, D_P, D_B) が与えられることとする。このとき、マクロブロック k に対する量子化パラメータ $Q(k)$ の算出を式 (7) により行う。

$$Q(k) = Mq(w(k) \times C_Q) \quad (7)$$

上式において、関数 Mq は量子化パラメータとして選択可能な値 (量子化スケールタイプにより決まる) のうち、引数に最も近いものを返し、 $w(k)$ はマクロブロック単位の視覚優先度パラメータを指す。 $w(k)$ は例えば、式 (8) (Test Model 5 の例) により、マクロブロック単位に予め算出されていることとする。同式において $act(k)$ は、当該マクロブロックにおける輝度ブロック分散の最小値を示し、 avg_act は $act(k)$ のピクチャ平均を示す。 C_Q はピクチャの目標量子化誤差電力を達成するという条件のもと設定される、ピクチャ内で共通的に基準とする量子化パラメータであり、ピクチャ単位に量子化誤差電力が目標値 D に近い値をとるように設定されることとする。具体的には、候補とする C_Q の範囲 (q_{min}, q_{max}) を式 (9) により限定した上で、 q_{max} を先頭に順次 d_q ずつ減じた値を候補として評価していくことで C_Q のサーチを行う。式中 qsc_{min}, qsc_{max} はそれぞれ、量子化スケールとして表現可能な最小値、最大値を指し、例えば、非線形の量子化スケールタイプ適用時には、それぞれ 1.0, 112.0 となる。本研究では、 d_q を便宜上 1.0 に設定する。候補 C_Q の評価値としては、基準となる量子化パラメータを C_Q とした際に得られるピクチャの量子化誤差電力を算出し、同評価値が目標値を下回る候補を検出した時点でサーチを終了し、同候補を C_Q として選択する。

$$w(k) = \frac{2 \times act(k) + avg_act}{act(k) + 2 \times avg_act} \quad (8)$$

$$\begin{cases} q_{min} = qsc_{min} / \max_k w(k) \\ q_{max} = qsc_{max} / \min_k w(k) \end{cases} \quad (9)$$

ここで選択された C_Q の大小により、現在選択されている量子化マトリクスの適性を判断できる点と、量子化マトリクスの更新がピクチャ単位に可能である点を考慮して、筆者らは先に、歪み最小化型ビット配分に基づく量子化マトリクスの更新手法を提案しており¹⁷⁾、本論文では同方式の併用を前提とする。

以上で述べたマクロブロックレイヤ処理の導入による効果は、従来方式が抱えていた仮想バッファによる不安定性を排除し、マクロブロック単位の優先度付けを忠実に反映可能な量子化制御の枠組みを提供できる点にある。事前実験の結果より、上述のマクロブロックレイヤ処理を非適用とし、これを Test Model 5 による仮想バッファ方式で置き換えた場合、ピクチャ符号化中の仮想バッファの変動は、結果として 0.75~1.25 の重み係数として、マクロブロック単位の量子化制御に反映されることを確認している。一方、マクロブロックレイヤ処理を適用した際には、方式の性質上、同問題は解消される。

3. 提案要素技術 一適応符号化制御の最適化一

3.1 マクロブロック単位量子化制御の高精度化

式 (8) に代表されるように、従来型のマクロブロック単位量子化制御は、基本的にマクロブロックの輝度分散のみを基準として視覚優先度パラメータを決定しており、人間の視覚感度を厳密に推定するものとはなっていない。この問題点を踏まえ、図 2 に示す階層構成をとるビット配分手法の提案を行う。図中、ピクチャレイヤ処理では 2.1 で述べた手法に沿ってピクチャ単位に目標量子化誤差電力を設定し、マクロブロックグループレイヤ (以下、MB グループと呼ぶ) 処理およびマクロブロック処理では、上述の視覚優先度パラメータ $w(k)$ を決定する。MB グループレイヤでは、画面をオブジェクト単位で扱う目的から、画面を構成する全マクロブロックに対し、グループ化に相当するクラスタリング処理を行う。提案するビット配分方式に関し、MB グループレイヤにおけるクラスタリング処理とマクロブロックレイヤにおける視覚優先度パラメータの算出に分けて、以下に述べる。

(1) MB グループレイヤ

クラスタリング処理で使用する分類基準値は、色差成分 Pb, Pr ならびに動き量から成る計 3 指標の組合せにより定義する。ここで便宜上、Pb, Pr はそれぞれ 5 段階に、動き量は 2 段階に分けた上で扱うことにより、最大で 50 種類の属性が存在することとする。Pb, Pr の評価にはマクロブロック平均値を使用し、動き量の評価に

は直前 (再生順) フレームを参照して得られた動きベクトル $V_f(k)$ に対し、カメラ操作 (パン, ズーム) に起因する成分を除去して得られるベクトル $V_g(k)$ の長さ $|V_g(k)|$ を使用することとした。カメラ操作に起因する動きベクトル成分の算出には、グローバル動きパラメータを使用した¹⁸⁾。以上のクラスタリング処理により分類された各 MB グループにおいて、その面積が小さい場合、分類基準値の近い隣接 MB グループ (隣接は上下左右の 4 方向とする) に統合する。統合処理は、すべての MB グループの面積が予め決められた閾値を上回るまで繰返し実行することとする。

次に、オブジェクト別の注視度を抽出する目的から、MB グループ単位に注視度パラメータ $v(j)$ を定義する。 j は MB グループの識別に用いる変数である。注視度パラメータの推定に際しては、以下の動視力特性および注視特性を考慮することとした。

- (I) 人間の動視力特性において、動きが小さい部分の方が良く見える傾向がある¹⁹⁾。
- (II) 人間は色の変化が大きい部分あるいはエッジ部分を注視する傾向がある²⁰⁾。

上記 (I),(II) の特性を注視度パラメータの算出に反映させるため、MB グループごとの動き量に関する指標と、周囲に対する変化度に関する指標を以下に定義する。

(I) 動き量に関する指標

まず MB グループの純粋な動き量を $L(j)$ とおき、式 (10) に従い算出する。式中 $l(l=0,1,\dots)$ は MB グループ j 内のマクロブロックの識別に用いられる変数であり、 $\bar{V}_f(j,l)$ は前述のベクトル $V_f(k)$ の長さ $|V_f(k)|$ に関する、MB グループ j 内のマクロブロック平均を示す。また、MB グループがある一定量 TH_L 以上の動きを含む際に、内部に含まれるマクロブロックごとの動きベクトルが散乱しているような場合には、注視度を低く設定すべきという配慮から、MB グループ内動きベクトルの散乱度を $R(j)$ とおき、式 (11) から式 (13) に従い算出することとした。式 (12) および式 (13) において、 $R(j,l,l')$ は MB グループ j に属するマクロブロック l の隣接マクロブロック (同一 MB グループに含まれる、 l' により識別) に対する $V_g(k)$ の散乱度を示す。式 (12) 中、 $(V_{gx}(l), V_{gy}(l))$ は前述のベクトル量 $V_g(k)$ の成分であり、 C_R は $R(j,l,l')$ に対して設ける上限値である。式 (13) 中 $\bar{R}(j,l,l')$ は $R(j,l,l')$ に関する MB グループ j 内の平均値を示す。

$$L(j) = \max(\bar{V}_f(j,l), 1) \quad (10)$$

$$R(j) = \bar{R}(j,l,l') \quad (11)$$

$$i) \quad L(j) > TH_L$$

$$R(j,l,l') = \min\left(\frac{(|V_{gx}(l)|+|V_{gx}(l')|)^2}{(V_{gx}(l)+V_{gx}(l'))^2+(V_{gy}(l)+V_{gy}(l'))^2}, C_R\right)$$

図 2 提案ビット配分方式の処理フロー
Proposed bit assignment process.

$$\text{ii) otherwise} \quad (12)$$

$$R(j, l, l') = 0 \quad (13)$$

(II) 周囲に対する変化度に関する指標

周囲に対する変化度を $K_b(j)$ とおき、式 (14) および式 (15) に従い算出する。式 (15) 中 $K_b(j, j', p)$ は、MB グループ j が異なる MB グループ (j' により識別) に隣接する、あるマクロブロック境界 (p により識別) において算出した変化度を示し、 $\overline{K_b}(j, j', p)$ は $K_b(j, j', p)$ に関する MB グループ j 内の平均値を示す。 w_v は動き量の変化度に与える重み係数である。また、同変化度が注視度に与える影響は、当該 MB グループと同様の属性を有するオブジェクトの発生頻度が小さいほど大きくなると考え、マクロブロックごとの分類基準値に関し、画面内のヒストグラム (該当するマクロブロック数) を求める関数 $freq$ を導入する。

$$K_b(j) = \overline{K_b}(j, j', p) \quad (14)$$

$$K_b(j, j', p) = w_v \times (V_g(j) - V_g(j'))^2 + (Pb(j) - Pb(j'))^2 + (Pr(j) - Pr(j'))^2 \quad (15)$$

以上で導入したパラメータおよび関数をもとに、式 (16) から式 (18) により注視度パラメータの算出を行うこととする。ただし、関数 Sig は引数の増加につれて、出力を徐々に漸近線に近づけるシグモイド関数である。指標 (I) に関しては、動き量 $L(j)$ および散乱度 $R(j)$ に対し、共に反比例となる特性をもたせることとした。指標 (II) に関しては、当該 MB グループの有する分類基準値の画面内ヒストグラムに反比例する関数と変化度 $K_b(j)$ の積を $K(j)$ とおき、式 (18) により定義した上で、 $K(j)$ が閾値を上回る場合に限り、 $K(j)$ に対し比例する特性を与えることとした。式 (18) 中 C_a , C_b は、画面内ヒストグラムに反比例する関数を定義するために用いる定数である。

$$\text{i) } K(j) > 1 \quad v(j) = Sig\left(\frac{K(j)}{L(j) \times R(j)}\right) \quad (16)$$

$$\text{ii) otherwise} \quad v(j) = Sig\left(\frac{1}{L(j) \times R(j)}\right) \quad (17)$$

$$K(j) = \left(\frac{C_a}{freq(V_g(j), Pb(j), Pr(j))} + C_b\right) \times K_b(j) \quad (18)$$

(2) マクロブロックレイヤ

マクロブロック視覚優先度パラメータ $w(k)$ の算出に関し、従来方式ではマクロブロックに閉じた属性のみを考慮し、式 (8) を用いて行っている。これに対し提案方式では、(1) で算出したオブジェクトの注視度パラメータ $v(j)$ と、当該マクロブロックが含むテクスチャの属性を示すパラメータ $t(k)$ (以下、テクスチャ属性パラメータ) をもとに式 (19) により算出することとする。ここで

$t(k)$ は、マクロブロック k において、量子化歪みがどの程度検知され易いかを示す指標に相当する。同指標の算出方法は、基本的に低周波数成分を多く含むほど大きな値を与える目的から、式 (8) の改良により以下に導入する。まずテクスチャ属性パラメータを、式 (8) に沿って $act(k)$ に基づき決定した際の問題として、精細なマクロブロック中に平坦なブロックが単独に存在する場合でも $act(k)$ が低い値となり、マクロブロック全体で量子化が精細に行われ、過度のビット配分を招く可能性がある。

$$w(k) = \frac{t(k)}{v(j)} \quad (19)$$

本問題を解決する目的から、視覚優先度パラメータの算出に使用する輝度分散については、当該ブロックおよび上下左右に隣接するブロックの計 5 ブロックの輝度分散値に rank order filter を適用し、2 番目に小さな値を当該ブロックの補正輝度分散として抽出する。ここで、マクロブロックを構成するブロックの補正輝度分散に関する最小値を $cact(k)$ とし、式 (8) にて $act(k)$ を $cact(k)$ で代用して得られる値を $cw(k)$ とおき、同値をマクロブロックのテクスチャ属性パラメータ $t(k)$ として用いることとする。 $t(k)$ の導出過程を、式 (20)、式 (21) にまとめる。 avg_cact は $cact(k)$ のピクチャ平均を示す。

$$t(k) = cw(k) \quad (20)$$

$$cw(k) = \frac{2 \times cact(k) + avg_cact}{cact(k) + 2 \times avg_cact} \quad (21)$$

ここで、MB グループ境界に位置するマクロブロック (以下、BMB と呼ぶ) では、隣接する MB グループの属性が混在している可能性が高く、加えて前述の注視特性 (II) に該当する可能性が高い。よって $t(k)$ の算出に関して、当該マクロブロックが BMB に該当する場合に限っては、例外的に式 (20) を適用せず、式 (22) を用いることで量子化制御上優遇することとした。同式において、 v_{max} および cw_{min} はともに、BMB および上下左右に隣接する計 5 マクロブロックより抽出するパラメータであり、それぞれオブジェクト注視度パラメータ $v(j_0)$ の最大値、 $cw(k)$ の最小値である。 j_0 は識別子 k で示されるマクロブロックが属する MB グループの識別子である。

$$t(k) = \frac{v(j_0)}{v_{max}} \times cw_{min} \quad (22)$$

3.2 ピクチャタイプの最適割当

前提として、 $M=3$ および $M=1$ の 2 パターンからの選択により、各種入力素材に対し、符号化効率を高めるという意味で適したピクチャタイプの選択が実現できることとする⁷⁾。以下では、歪み最小化型ビット配分の拡張に基づき、1 パスで $M=3/M=1$ の適応選択を高精度に実現する方式を提案する。ここで、1 パスで適応選択が可

能であるとは、本線の符号化とは別系統で事前符号化を行うことなく、一般的なエンコーダシステムと同様に、当該フレームの前処理および近傍フレームの符号化結果をもとに、1系統で適応制御を実現できることを意味し、放送での適用を想定する場合には必須である。

まずMPEG-2符号化においては、ピクチャタイプとしてI,P,Bの3種類の併用が可能である。ピクチャ単位の適応制御による効果が高いと考えられるものにシーンチェンジ点が挙げられ、シーンチェンジ点での画質劣化を防止する目的から、割当ビット数の算出やIピクチャの挿入タイミングに関する対策を検討した例がみられる¹³⁾。本検討においてもIピクチャの挿入は、シーンチェンジ点にリンクして行うこととし、符号化に先立ち当該ピクチャのシーンチェンジ情報が得られるという前提を設ける。ただし、本研究の対象が放送であることから、受信機でのチャンネル切替を考慮し、定期的なIピクチャの挿入を義務付ける(例えば、0.5秒に1回)。また前提として、シーンチェンジ点においては、初期状態としてM=3を選択し、以降はIピクチャおよびPピクチャの挿入を行うタイミングに合わせ(シーンチェンジ点直後のPピクチャを除く)、更新の判定を行うこととする。M=3/M=1適応選択の一例を図3に示す。更新タイミングにおいてM=3/M=1の選択に使用する判定指標(以下、単に判定指標と呼ぶ)を以下に導出する。

更新タイミングでは、M=3選択時、M=1選択時の双方について、直後に符号化すべきPピクチャに、歪み最小化型ビット配分を適用し、目標量子化誤差(式(1)中 D_P)を算出し、それぞれ $D_{P,M3}$ 、 $D_{P,M1}$ とおく。ここでPピクチャ符号化の観点から、 $D_{P,M3}$ が $D_{P,M1}$ を上回る場合に限りM=1を適用することとする。ただし $D_{P,M3}$ 、 $D_{P,M1}$ の算出を行う上で、それぞれM=3、M=1の双方に対し、以下のパラメータが必要となる。

- ・ピクチャあたりのサイド情報ビット数(ピクチャタイプごと)
- ・量子化器入力信号の分散： $(\sigma_{xI}^2, \sigma_{xP}^2, \sigma_{xB}^2)$
- ・R-D関数： (f_I, f_P, f_B)

このうち選択中のM値に関しては、近傍の符号化結果より上記を取得すれば良いが、非選択中のM値に関し

ては過去の符号化結果を用いた推定により算出する。まずサイド情報ビット数に関しては、文献⁷⁾の手法により推定することとした。パラメータ $(\sigma_{xI}^2, \sigma_{xP}^2, \sigma_{xB}^2)$ と関数 (f_I, f_P, f_B) に関しては、同一ピクチャタイプを有する符号化済ピクチャのうち再生時刻が当該ピクチャに最も近いものの符号化結果より算出する。ただし、Pピクチャに対する σ_{xP}^2 の算出に関してはM=3とM=1で予測フレーム間隔が異なるため、次式による補正を行う。

i) M=3選択時

$$\sigma_{xP}^2 = \sigma_{xP,M3}^2 \times \frac{1}{W_E} \quad (23)$$

$\sigma_{xP,M3}^2$ は、近傍のPピクチャ符号化(予測フレーム間隔=3)において取得した σ_{xP}^2 を指す。

ii) M=1選択時

$$\sigma_{xP}^2 = \sigma_{xP,M1}^2 \times W_E \quad (24)$$

$\sigma_{xP,M1}^2$ は、近傍のPピクチャ符号化(予測フレーム間隔=1)において取得した σ_{xP}^2 を指す。

W_E はM=3とM=1でPピクチャの予測フレーム間隔が異なる点を考慮して導入する重み係数であり、予測フレーム間隔の増加に伴ってPピクチャの予測効率が低下する度合を反映する。 W_E は文献⁷⁾の手法により算出することとした。算出手順を以下に説明する。

まず、予測効率の低下度合いを推定する際に、利用すべき情報を選定するにあたり、現在選択中のM値に依らず、共通的に参照できるものであることが制限事項として挙げられる。そこで、1フレーム間隔の順方向予測結果を採用し、M=1を選択中であれば近傍Pピクチャでの予測結果を、M=3を選択中であれば近傍Bピクチャでの予測結果を参照することとした。手順としては、1フレーム間隔の順方向予測誤差信号の算出において、トップフィールドとボトムフィールドについて参照フィールドを限定しない場合の誤差2乗和を計算し、それぞれ E_{top} 、 E_{btm} とする。加えて、トップフィールドとボトムフィールドについて参照可能なフィールドを、予測距離が大きい一方に限定した場合の誤差2乗和を計算し、それぞれ E'_{top} 、 E'_{btm} とする(図4参照)。このとき次式により W_E を決定する。以上の算出方式は、Pフレーム間隔が2フレーム離れることによる予測性能の低下を厳密に推定するものではないが、事前実験の結果より、実測値と推定値の間に0.9を上回る高い相関を確認している。

$$W_E = \min\left(\frac{E'_{top}}{E_{top}}, \frac{E'_{btm}}{E_{btm}}\right) \quad (25)$$

4. 性能評価

3で導入した提案要素技術の性能評価を行う目的から符号化実験を行った。以下で共通に使用する符号化シミュレーション条件を表1に示す。事前の実験結果より、素材および符号化レートに依らず、式(2)の γ_x^2 は1.0に設定し、式(3)の w_B は1.4に設定した。

図3 M=3/M=1適応選択の一例
Example of adaptive selection from M=3 and M=1.

図 4 重み係数 W_E の算出
Calculation of the weighting factor W_E .

表 1 共通の符号化条件
Condition of coding experiments.

符号化方式	MPEG-2 MP@HL(4:2:0)
符号化レート	11Mbps, 15Mbps
GOP 構造	N: 15 M: M=3 または M=1
動き補償	フレーム予測/フィールド予測適応
動き検出	± 63.5 画素×± 31.5 ライン, 半画素 (1frame 間隔に対して)
DCT タイプ	フレーム DCT/フィールド DCT 適応
量子化制御	量子化スケールタイプ: 非線形

(1) マクロブロック単位量子化制御の高精度化

3.1 で導入した提案方式の適用により, どの程度の画質向上が見込めるかを検証する目的から, マクロブロックごとの優先度付けを, より厳密に行った際の画質改善を評価した. 歪み最小化型ビット配分の制御下で, 3.1 で提案した視覚優先度パラメータの割当方式を組込んだ上で, 符号化実験を行った. 事前の実験結果より, 式 (12) における TH_L , C_R , 式 (15) における w_v , 式 (16) および式 (16) における関数 Sig , 式 (18) における (C_a, C_b) については次式の設定を用いた.

$$\left\{ \begin{array}{lcl} TH_L & = & 3.0 \\ C_R & = & 10.0 \\ w_v & = & 0.1 \\ Sig(x) & = & \frac{4.0}{exp(0.5-x) \times 2.0 + 1.0} \\ (C_a, C_b) & = & (100.0, 0.1) \end{array} \right. \quad (26)$$

一例として, ITE HDTV 標準テスト画像 "Sprinkling" を入力とし, 15Mbps の符号化レートで得られた再生画像を図 5 に示す. 事前の実験結果より M 値は最良値として M=1 に設定した. 比較方式として, 同一の符号化条件下で 2.2 に相当する方式を適用した際の結果を同図中 "Conventional" として示す. 比較方式において視覚優先度パラメータの算出は式 (8) に準じて行った. ただし, 再生画像は画質の違いが顕著な部分として, 人物の顔周辺領域のみを切り出して示した. 結果より, マクロブロックごとの優先度付けの最適化による主観画質の向上が見られ, 特に顔の平坦部分におけるブロック歪みの低

(a) Original picture

Conventional Proposed
(b) Coded picture

図 5 再生画質の比較
Comparison of reconstructed picture quality.

表 2 使用したテスト画像
Selected test sequences.

Scene No.	Seq. No.	使用 frame	シーン特性
1	16	#30 - #59	横バンの動き最大部
2	16	#84 - #113	シャチのジャンプ
3	19	#0 - #29	入場行進
4	23	#0 - #29	緩やかな前後の動き
5	25	#48 - #77	女性の立ち上がるシーン
6	7	#82 - #11	路面バスが前面で動く (難易度: 高)
7	10	#84 - #113	車, バスの往来
8	30	#50 - #79	行き交う通行人
9	34	#45 - #74	選手を追うカメラ操作を含む
10	39	#44 - #73	アナウンサーの手の動くシーン

減, ならびに目やスカーフなど各種輪郭の忠実な再現といった, 改善効果が見られる. その他の素材に対する実験結果も踏まえ, 提案方式の導入による主観画質の向上は, 方式の性質上, 一般に以下の条件を満たす素材において特に顕著であることを確認した.

- ・画面を構成するオブジェクトが, 動き量と色差により特徴づけられる.

- ・実際の注視領域が特定オブジェクトに偏って存在し, 周辺オブジェクトに対する注視度の差が激しい.

(2) ピクチャタイプの最適割当

まず 3.2 で提案した判定指標の算出精度を評価する目的から, 符号化シミュレーション実験により, $D_{P,M3}$ および $D_{P,M1}$ の算出を行った. 入力には, ITE HDTV 標準テスト画像より計 10 シーン (シーン長は 30frame) を選定して使用した. 使用したテスト画像のシーケンス番号, 使用 frame 番号, シーン特性を表 2 に示す. GOP 構造を除き符号化条件は表 1 に従うこととした. 各種テ

図 6 $D_{P,M3}$ の算出結果
Observed value of $D_{P,M3}$.

図 7 $D_{P,M1}$ の算出結果
Observed value of $D_{P,M1}$.

スト画像入力に対し、 $M=1$ 選択時に算出した $D_{P,M3}$ の推定値、 $M=3$ 選択時に算出した $D_{P,M1}$ の推定値の結果をそれぞれ図 6、図 7 に示す。符号化レートは、15Mbps および 11Mbps の 2 通りに設定した。図 6 および 図 7 では、比較のため、推定対象とするパラメータを、非選択中の M 値による符号化結果より取得した実データを重ねて示した。結果より、 $D_{P,M3}$ 、 $D_{P,M1}$ のいずれについても、提案した算出方法により、符号化レートに依存せず、高い精度で推定可能であることがわかる。

表 3 ビクチャタイプの適応選択による PSNR の結果 (15Mbps)
Results of PSNR obtained by the scene adaptive picture type assignment at the bit-rate of 15Mbps.

テスト画像	PSNR dB (M=3 選択率)			
	Proposed	ref. ⁷⁾	M=3 固定時	
			Total	B 以外
Whale Show 1	27.93(0)	27.93(0)	26.94	26.99
Whale Show 2	25.24(0)	24.72(0)	24.67	25.01
Opening Ceremony	26.24(100)	26.24(100)	26.24	26.38
Green Leaves	23.46(100)	23.46(100)	23.46	23.55
Japanese Room	37.94(0)	37.39(61)	37.01	37.69
European Market	33.98(0)	33.98(0)	32.68	32.92
Streetcar	34.52(0)	34.39(35)	34.23	34.64
Crowded Crosswalk	25.32(0)	25.32(0)	33.47	34.61
Ice Hockey	37.28(0)	36.74(100)	36.74	37.5
Weather Report	38.93(31)	38.82(63)	38.85	39.44

表 4 ビクチャタイプの適応選択による PSNR の結果 (11Mbps)
Results of PSNR obtained by the scene adaptive picture type assignment at the bit-rate of 11Mbps.

テスト画像	PSNR dB (M=3 選択率)			
	Proposed	ref. ⁷⁾	M=3 固定時	
			Total	B 以外
Whale Show 1	26.11(0)	26.11(0)	25.86	25.63
Whale Show 2	24.14(0)	24.14(0)	23.67	23.66
Opening Ceremony	25.55(0)	25.55(0)	25.04	24.91
Green Leaves	21.87(0)	21.72(100)	21.72	21.49
Japanese Room	36.40(0)	36.40(0)	34.91	34.79
European Market	31.48(0)	31.48(0)	29.92	29.36
Streetcar	33.20(0)	33.20(0)	32.19	31.99
Crowded Crosswalk	33.46(0)	33.46(0)	31.04	31.38
Ice Hockey	36.38(0)	36.38(0)	35.26	35.17
Weather Report	37.15(0)	37.13(80)	37.01	37.4

次に判定指標を用いた適応選択手法を評価する目的から、符号化実験を行った。入力には表 2 で選定した 10 カットをランダムに連続再生することで 300 フレームの素材を用意した。ビクチャタイプ割当に関する提案手法の単独効果を評価する目的から、マクロブロック単位量子化制御について、3.1 の提案方式は非適用とした。表 3 に 15Mbps の符号化レートにおいて、提案したビクチャタイプの適応選択を適用した際の PSNR の結果、ならびに更新タイミングにおける $M=3$ の選択率をカットごとに示した。比較のため、文献⁷⁾による $M=3/M=1$ の適応選択方式、ならびに $M=3$ 固定割当による結果も合わせて記載した。 $M=3$ 固定割当の符号化結果については、B ビクチャを除くビクチャに限定した平均 PSNR と全体の PSNR に分けて示した。11Mbps の符号化レートにおいて同様に取得した結果を表 4 に示す。

結果より、一部の素材において文献⁷⁾による手法と比較し性能改善が確認できる。また提案方式の導入により、文献⁷⁾適用時および固定割当時に対する効率低下は認められていない。11Mbps の符号化レートでは、いずれの適応選択方式においても、 $M=1$ を選択するケースが大半を占める理由から、提案方式の導入による効果は 15Mbps の符号化レートにおいて顕著である。

文献⁷⁾に対するゲインについて以下に考察する。文

献⁷⁾では、BピクチャのDCT係数部分に割当てられるビット数を0と仮定した上で、 $M=3/M=1$ の双方におけるサイド情報ビット数と符号化レートのみを考慮し、当該PピクチャのDCT係数部分に割当可能なビット数を推定し、判定指標として参照している。これに対し提案方式では、 $M=3/M=1$ の双方につき、歪み最小化型ビット配分の枠組を利用して、当該Pピクチャの符号化歪みを算出し、判定指標として参照することで、より高精度な適応選択を実現するものである。以上の相違点に起因して、提案方式の導入により、文献⁷⁾に対する性能改善が実現されたものと結論付けられる。

5. む す び

デジタル放送への応用を目的とし、低ビットレートHDTV符号化の高画質化アプローチとして、MPEG-2適応符号化制御の最適化手法を提案した。提案手法としては、歪み最小化型ビット配分に基づき実現される、マクロブロック単位量子化制御の高精度化、およびピクチャタイプ割当の最適化につき検討を行った。符号化実験結果より、提案方式の導入による効果として、高難易度素材に対する主観画質の向上、および符号化効率の改善を確認した。

本論文で扱った最適化対象は、MPEG-2の符号化シンタックス上考えられる、適応符号化制御のごく一部に閉じたものであり、例えば、動きベクトル検出や符号化モード判定など、最適化対象を広げ、方式単体のゲインを積み上げていくことで、低ビットレートでのHDTV放送品質確保が、より確実なものになると期待され、さらなる検討課題としたい。

本研究は、通信・放送機構からの委託研究「デジタル放送用HDTV高圧縮技術に関する研究開発」に基づき行われたものである。

日頃ご指導頂く株式会社KDDI研究所代表取締役所長浅見 徹氏、ならびに同所代表取締役副所長 松島 裕一氏に感謝いたします。

〔文 献〕

- 1) 黒住, 神田, 杉本, 中須, 大塚, "画像構造適応選択による MPEG-2 低ビットレート符号化の検討", 映像学技報, BCS2001-12, pp.13-18(March 2001)
- 2) 甲藤, 太田, "MPEG 圧縮効率の理論解析とその符号量制御への応用", 信学技報, IE-95-10(April 1995)
- 3) 水野, 鹿喰, 野尻, 大塚, 田中, "ピクチャ別レートひずみ特性に基づいた符号量配分手法の検討", 信学論, J82-D-II, pp.1239-1251(Aug. 1999)
- 4) J. Lee and B. W. Dickinson, "Rate-Distortion Optimized Frame Type Selection for MPEG Encoding", IEEE Trans. Circuits and Systems for Video Technology, 7, 3, pp.501-509(1997)
- 5) 米山, 中島, 柳原, 菅野, "シーンアダプティブな可変 GOP 構造を持つ MPEG 符号化アルゴリズム", 1999 信学全大, D-11-10(March 1999)
- 6) 横山, "MPEG-2 ビデオ符号化における適応 GOP 構造選択", 2000 信学全大, D-11-40(March 2000)
- 7) 内藤, 小池, 和田, 松本, "MPEG-2 符号化構造の最適化に基づく低レート HDTV 符号化方式", 映像学誌, 56, 2, pp.253-262(Feb. 2002)
- 8) H. Sun, W. Kwok, M. Chien, C. H. John, "MPEG Coding Performance Improvement by Jointly Optimizing Coding Mode Decisions and Rate Control", IEEE Trans. Circuits and Systems for Video

Technology, 7, 3, pp.449-458(1997)

- 9) M. C. Chen, A. N. Willson, "Motion-Vector Optimization of Control Grid Interpolation and Overlapped Block Motion Compensation using Iterated Dynamic Programming", IEEE Transactions on Image Processing, 9, 7, pp.1145-1157(2000)
- 10) 青木, 浜田, 小西, 水野, 神田, 中須, 湯山, "MPEG2 映像符号化の画質改善の基礎検討", 映像学技報, BCS97-41, pp.49-54(May 1997)
- 11) 大塚, 黒住, 神田, 西田, 中須, "動きベクトル情報量低減による MP@HL 符号化の画質改善の検討", 映像学技報, BCS98-1, pp.1-6(March 1998)
- 12) 内藤, 和田, 松本, 羽鳥, "ビット配分の最適化に基づく放送用 MPEG-2 符号化制御の検討", 映像学技報, BCS2002-1(Feb. 2002)
- 13) M. Lee, S. Kwon, J. Kim, "A SCENE ADAPTIVE BITRATE CONTROL METHOD IN MPEG VIDEO CODING", VCIP, SPIE Vol. 3024, pp.1406-1416(1997)
- 14) Test model Editing Committee, "Test Model 5", ISO/IEC JTC1/SC29/WG11 N0400(1993)
- 15) 滝嶋, 和田, "蓄積伝送型ビデオコーデックにおける符号化制御方式", 映像学誌, 47, 10, pp.1304-1312(Oct. 1993)
- 16) G. Keesman et al., "Bit-rate control for MPEG encoders", Signal Processing: Image Communication vol.6, pp.545-560(1995)
- 17) 内藤, 小池, 和田, 松本, "歪み最小化型ビット配分の高度利用に基づくデジタル放送用 MPEG-2 符号化制御の最適化", 映像学技報, BCS2002-38(Oct. 2002)
- 18) 上倉, 渡辺, "動画像符号化におけるグローバル動き補償法", 信学論, 76-B-I, 12, pp.944-952(Dec. 1993)
- 19) 高木, 小池, 和田, 松本, "次フレームでの符号化参照の有無による量子化制御方式", 信学ソ大, D-11-5(Sep. 2000)
- 20) 田中, 井上, 中津, "誘目性に寄与する物理的特徴量を基にした画像注目領域の注出", 映像学誌, 52, 6, pp.881-890(June 1998)

内藤 直人 1994 年, 早稲田大学理工学部電子通信学科卒業。1996 年, 同大学大学院修士課程修了。同年, 国際電信電話 (株) に入社。現在, (株)KDDI 研究所にて, 主に HDTV, SDTV の動画像符号化および伝送方式の研究・開発に従事。本学会, 鈴木記念賞, 研究奨励賞を受賞。正会員。

小池 恭一 1983 年, 新潟大学工学部情報工学科卒業。1985 年, 東北大学大学院修士課程修了。同年, 国際電信電話 (株) に入社。以来, テレビ会議・電話システム, 動画像の高効率符号化方式の研究・開発に従事。現在, (株)KDDI 研究所画像通信グループリーダー。本学会, 鈴木記念賞を受賞。工学博士。正会員。

松本 修一 1979 年, 北海道大学大学院修士課程修了。同年, 国際電信電話 (株) に入社。以来, 研究所にて, HDTV, SDTV の高効率符号化方式の研究・開発に従事。現在, (株)KDDI 研究所取締役。本学会, 技術振興賞開発賞, 丹羽高柳賞著述賞を受賞。工学博士。正会員。

羽鳥 好律 1971 年, 東京大学工学部電気工学科卒業。同年, 国際電信電話 (株) に入社。2003 年より, 東京工業大学大学院総合理工学研究科教授。画像信号の高効率符号化方式, ネットワークインターフェースの研究に従事。本学会, 鈴木記念賞, 丹羽高柳賞論文賞, 業績賞を受賞。工学博士。正会員。