

レート制御可能な動画の準可逆符号化

Near-Lossless Video Coding with Rate Control Ability

横田 達朗*

青森 久*

松田 一郎*

伊東 晋*

Tatsuro YOKOTA

Hisashi AOMORI

Ichiro MATSUDA

Susumu ITOH

1. はじめに

先に我々は、ブロック適応3次元予測を用いた動画の準可逆符号化方式を提案した [1]。この方式は、再生誤差の絶対値が許容値 d 以下となることを保証しつつ、可逆符号化の場合よりも大幅に高い圧縮率を達成可能であり、素材映像など高品質な動画データを保存する用途に適している。しかし、許容誤差 d を変更する以外に符号化レートを制御する手段がなく、指定レートでの映像伝送を行うことは困難である。一方、文献 [2, 3] では画面内に2種類の許容誤差を混在させることで静止画像の符号化レートを任意の値に一致させる手法が提案されている。本稿では、この手法を動画の準可逆符号化用に拡張し、フレームまたは GOP 単位で符号化レートを指定可能とする方式について検討する。

2. ブロック適応3次元予測を用いた準可逆符号化

本方式では、フレーム毎に複数の線形予測器を用意し、これらを 8×8 画素のブロック単位で適応選択している [4]。符号化区間 (GOP) の先頭フレーム (I ピクチャ) では、静止画像の場合と同様な2次元予測 [3] が適用される。一方、2枚目以降のフレーム (P ピクチャ) に関しては、図1に示すように当該フレームの符号化済み近傍画素群 $\{p_k\}$ だけではなく、前フレーム上の画素群 $\{q_k\}$ も同時に参照する3次元予測を採用している。注目画素 p_0 が属するブロックに m 番目の予測器が割り当てられた時の予測値 $\hat{s}_t(p_0)$ は次式で表せる。

$$\hat{s}_t(p_0) = \sum_{k=1}^{K_p} a_m(k) \cdot \tilde{s}_t(p_k) + \sum_{k=1}^{K_q} a_m(K_p + k) \cdot \tilde{s}_{t-1}(q_{k-1} + v) \quad (1)$$

ここで、 K_p, K_q は予測に用いる画素 p_k および q_k の数 (予測次数)、 $a_m(k) (k = 1, 2, \dots, K_p + K_q)$ は各画素に対する予測の重み (予測係数)、 $\tilde{s}_t(p_k), \tilde{s}_{t-1}(q_k)$ は当該フレームおよび前フレームの再生値をそれぞれ表している。また、 v は、四分木分割に基づいた可変サイズのブロック単位で検出した整数画素精度の動ベクトルである。(1)式に基づいて算出された各画素の予測誤差 $e(p_0) = s_t(p_0) - \hat{s}_t(p_0)$ は、許容誤差に対応したステップ幅 $\Delta e = 2d + 1$ の一様量子化器で量子化された後、次

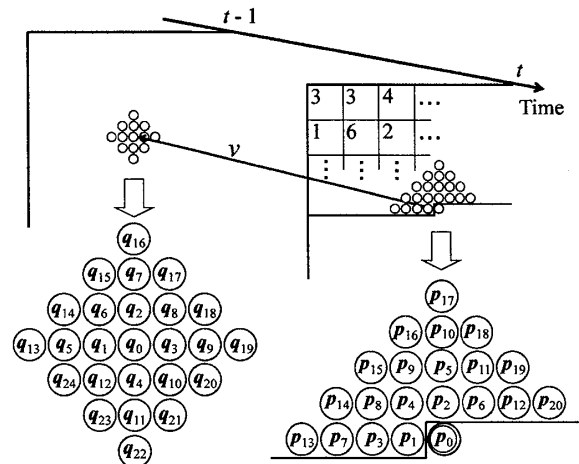


図1 ブロック適応3次元予測 (P ピクチャ)

式の特徴量 U に基づいたコンテキストモデリングにより適応的に算術符号化される。

$$U = \sum_{k=1}^6 |e(p_k)| + \sum_{k=0}^4 |e(q_k)| \quad (2)$$

なお、予測係数 $a_m(k)$ やブロック毎に割り当てられる予測器の選択状況 m 、動ベクトル v といったパラメータは、可逆符号化 ($d = 0$) を前提として見積もられた符号量が最小となるよう予めフレーム毎に最適化され、付加情報としてビットストリームに多重化される [1]。

3. レート制御可能な準可逆符号化

上記の符号化方式において、フレーム毎に許容誤差 d の値を変えると画質および発生符号量も変化するが、その変動幅が大きいため、正確なレート制御を行うことが難しい。そこで、文献 [2, 3] と同様に2種類の許容誤差を画素毎に切り替えることで、フレーム毎の符号化レートを指定した値に一致させる手法を導入する。まず、フレーム全体で許容誤差を $d = 0, 1, 2, \dots$ と変えながら仮の符号化処理を実行し、指定された符号化レートを上回る最大の d の値を決定する。原理的には、 d および $d+1$ が割り当てられる画素の比率を調整することで指定レートでの符号化が可能となるが、視覚的には異なる許容誤差で符号化された画素の分布が検知されないことが望ましい。そこで、人間の視覚感度が輝度変化の激しい領域で低下すること (マスキング効果) に着目し、平坦部以外に優先的に大きい許容誤差を割り当てるものとする。具

*東京理科大学 理工学部 電気電子情報工学科

体的には、コンテキストモデリングにおいて予測誤差のアクティビティの指標として用いた特徴量 U を利用し、その値がしきい値 R_{TH} を超える画素に対して許容誤差 $d+1$ を割り当てる。当該フレームの発生符号量はしきい値 R_{TH} に対して単調増加と見なせるため、二分法と同様の手法により指定レートを実現する R_{TH} の値を効率よく求めることが可能である。

4. GOP の符号量を一定とするレート制御手法

動画像符号化において、各フレームに一定の符号量を割り当てるのが画質の点で最良とは限らない。そこで、GOP 全体の符号量が一定という条件の下で各フレームの性質(符号化難易度)に応じたビット配分を行い、平均画質の向上を図る。本稿では、 i 番目のフレームの符号化難易度を表す指標として、(1)式において予測次数を $K_p = 2, K_q = 1$ (Iピクチャでは $K_p = 2, K_q = 0$) として設計した簡易な線形予測器による予測誤差のエントロピー H_i を採用した。ここで、GOP 内の目標符号量を T [bits] としたとき、各フレームに割り当てられる符号量 T_i は次式によって与えられる。

$$T_i = \left(T - \sum_{k=1}^{i-1} \hat{T}_k \right) \times \frac{H_i}{\sum_{k=i}^N H_k} \quad (3)$$

ただし、 N は GOP 内のフレーム数、 \hat{T}_k は符号化が終了したフレームにおける実際の符号量を表している。

5. 特性評価とまとめ

ITE/ARIB 標準動画像 version 2 の Y 信号を縦横共に 1/4 に縮小した 8bit 精度のモノクロ動画像 (480 × 270 画素, 25 フレーム分, 30 Hz) を対象として符号化シミュレーションを実施した。3次元予測器の予測次数は、 $K_p = 20, K_q = 25$ (Iピクチャでは $K_p = 30, K_q = 0$) とそれぞれ設定した。表1は、目標符号化レートを 1.200 [bits/pel] と指定して、フレーム単位および GOP 単位のレート制御を行った場合の、GOP 内の平均 SNR (全フレームの平均 2 乗誤差より算出) を調査した結果であ

表 1 GOP 内の平均 SNR [dB]

Image	フレーム単位	GOP 単位
<i>Flash Photography</i>	50.70	51.26
<i>Fountain (follow)</i>	46.58	48.22
<i>Horse Racing (dirt)</i>	44.24	45.88
<i>Red Leaves (pan up)</i>	44.13	45.30
<i>Studio Concert (confetti)</i>	45.76	46.32
<i>Truck Train</i>	49.37	50.23

る。いずれの場合も、指定レートに対して 0.02% 以内の精度で実際の符号化レートを一致させることが可能であることを確認している。これより、GOP 単位でレート制御を行った方が、フレーム単位で一定レートを割り当てる場合より、平均 SNR の点で有利であることがわかる。図2は、画像 *Flash Photography* におけるフレーム毎の SNR の推移を表したものである。この画像はフラッシュを炊きながらモデルを写真撮影するシーンであり、フラッシュの点灯により輝度レベルが急激に変化するため、符号化難易度も大きく変動する傾向がある。このためフレーム単位のレート制御を行うと SNR が大きく低下する場合がありますのに対し、GOP 単位のレート制御では SNR の変動が抑えられ、結果的に画質向上につながっていると考えられる。

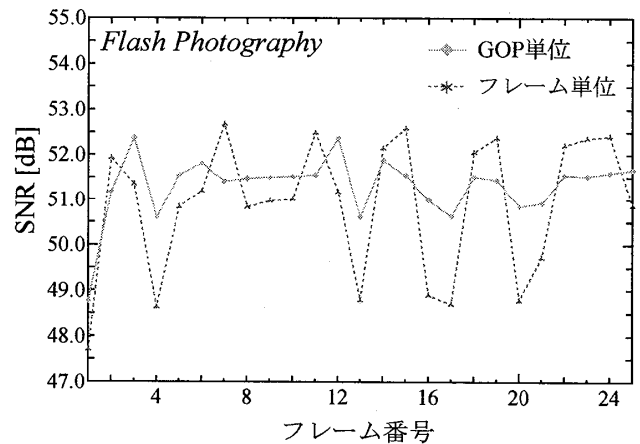


図 2 フレーム毎の SNR の推移

【参考文献】

- [1] 三浦智洋, 高橋一哉, 松田一朗, 伊東晋: “動き補償と適応 3 次元予測に基づいた動画像の準可逆符号化”, 信学総大, No.D-11-35, 2007
- [2] J. Jiang et al.: “A low cost design of rate controlled JPEG-LS near lossless image compression”, Image Vision Comput, Vol.19, No.3, 2001
- [3] 坪井宣親, 宮本駿昂, 青森久, 松田一朗, 伊東晋: “レート制御可能な静止画像の準可逆符号化方式”, FIT2009, Vol.3, No.I-043, 2009
- [4] 塩寺太郎, 梅津有司, 前田弘樹, 松田一朗, 森岡一幸, 伊東晋: “フレーム毎に MC と 3 次元予測を最適化する動画像の可逆符号化方式”, 映情学誌, Vol.60, No.7, pp.1051-1058, 2006