

Padding法を用いた任意形状直交変換符号化方式の一検討

高木 幸一 小池 淳 松本 修一

(株)KDD 研究所 画像通信グループ

〒 356-8502 埼玉県上福岡市大原 2-1-15

E-mail : ko-takagi@kddlabs.co.jp

あらまし 本稿では、任意形状直交変換のための Padding 法について検討したので報告する。MPEG-4 では、従来の矩形単位の直交変換の弊害を避けるために画素埋め込み (Padding) 等による可変形状領域に対する直交変換符号化が採用されている。この Padding に対して筆者らはエントロピーと量子化誤差を評価する関数を設定し、両者最小の規範に基づいて最適化する手法を提案する。さらに、テスト画像を用いて計算機シミュレーションを行い、本手法の有効性を示す。

A Study of Image Coding Based on Arbitrarily-shaped Orthonormal Transform Using Padding Method

Koh'ichi TAKAGI

Atsushi KOIKE

Shuichi MATSUMOTO

KDD R&D Laboratories

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

abstract In this paper, we have studied the padding method for arbitrarily-shaped orthonormal transform. In order to avoid the bad influence due to conventional orthonormal transform based on rectangle, it is adopted arbitrarily-shaped image coding, for example, padding method, etc. in MPEG-4 verification model (VM). We defined an evaluation function approximating their quantized error and entropy, and proposed how to minimize both of them. As a result of the computer simulation for 8x8 block texture coding, the proposed method is performed on the average 4 or 5 dB greater than MPEG-4 VM padding method.

1 はじめに

現在、国際標準となっている映像信号の圧縮手法において矩形単位の2次元DCTがよく用いられている。この方法は平坦な画像ならば非常に効率良く符号化できるが、エッジなど高周波成分を含む矩形に関してはエネルギーが分散してしまい、逆に符号化効率を悪化させてしまう[3]。そこで、このような弊害を防ぐ方法として画像を均一な性質を有する領域に分割し符号化する領域分割符号化が注目を集めている。MPEG-4においても、物体(content)ごとに独立に符号化する方法(content-based coding)が採用されており、この類の弊害を低減させている。

具体的に、任意形状領域符号化のための直交変換の方法として、画像の相関モデルから変換基底を導出する方法[7]、基底は従来のDCT(など)を用い、画像信号を表現する方法[8](つまり、復号が従来通りできる)など様々な方法が提案されている。また、MPEG-4 VM12[10]でもSA-DCT[6]やpadding法が採用されている。

この中でpadding法とは、符号化すべき物体の外側(すなわち背景領域)に何らかの画素値を埋め込み、従来の矩形単位のDCTを適用する方法である。すなわち、復号が従来通りでき、非常に簡単であるため、この埋め込む画素値を如何に決定するかについての研究が数多くされてきている。そこで本稿でもこのpadding法に着目することとする。

ところで、従来のpadding法として、単純に画素値の範囲から求められた平均値を埋め込む方法や物体領域の画素値の平均値を埋め込むMean Replacement Padding(MRP)法[4]、1/2,1/4,1/8各対称点にある(物体領域の)画素値を埋め込む2D-DCTの特徴に特化した方法(Symmetric Padding)[9]などが提案されている。また、MPEG-4 VM12では、textureを符号化するためのpadding法としてMRP法にLow Pass Filteringを組み合わせたLow Pass Extrapolation(LPE)法が、動き推定、動き補償をする際のpadding法として物体境界の画素値の平均値を埋め込む(Macro-)Block Repetitive Padding法が用いられている[10]。しかし、上述の手法は主に処理速度を重視したものであり、ビットレート-歪み特性の立場からの検討はこれまでほとんどない。

そこで筆者らは、量子化誤差およびエントロピーを近似する関数を設定しそれらが最小の規範のもとで、埋め込むべき画素値を決定する手法を提案する。その手法を2に、 8×8 の矩形領域に適用した結果を3に示す。さらに、本手法をIntra Pictureに適用するために、画像を $2N \times 2N$ 画素の矩形領域ごとに分け、それをさらに4個に分割して符号量、およびSNRの関係から準最適な符号化をした。その方法を4に、その結果を5に示す。そして、最後

に本稿をまとめる。

2 padding 法

本章では、本稿で提案する準最適padding法について述べる。

任意の 8×8 画素行列 $X = \{x_i | 0 \leq i \leq 63\}$ にDCTを適用してできた行列を $Y = \{y_i | 0 \leq i \leq 63\}$ とする(簡単のため1次元で書く)。ここでは、 X はすでに分割されており、物体領域(画素値が設定されている領域) X_o 、背景領域(画素値を埋め込む領域) X_b から成るとする。そして、 X_b の要素を未知数(のベクトル) $z = (z_0, z_1, \dots, z_{N-1})^T$ (すなわち背景領域の要素数 $|X_b| = N$ 個)と置く(図1)。

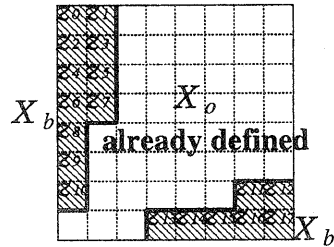


図1: 領域分割された 8×8 行列

このとき、 z を如何に置けば最もDCTが効率良くできるかを考えたい。 Y は X にDCTを適用した結果なので、 y_i は x_i の線形結合で、さらには z_i の線形結合で書き表すことができる。そこで、 $a_i = (a_0, a_1, \dots, a_{N-1})$ (a_i はDCT係数の計算で定まる定数)として

$$y_i = a_i z + a_N \quad (1)$$

と記述できる。

ところで、DCTに代表される直交変換を用いた符号化を効率良くするためには以下の条件を満たしていることが望ましい。

- 決められた量子化器に対し、量子化誤差が小さい。
- DCT係数の分散が大きく、低域にパワーが集中する。

そこで、本稿では以上の条件に見合う評価式を設定し、その式を最適にすることで z を逆算する方式を提案する。

以下にその評価関数を述べる。なお、下記の関数において x_i, y_i は z の関数となっていることに注意する。

2.1 量子化誤差見積もり関数の設定

「量子化誤差」を見積もる関数を設定する。本来ならば

$$E_q(z) = \frac{1}{64} \sum_i (y_i \text{の量子化誤差})^2 \quad (2)$$

とすべきである。しかし、これを解析的にやるため、評価関数は1次微分可能な関数とした。本稿では線形量子化を用いることとして、量子化ステップサイズを Δ とし、量子化誤差の大きさを評価する式を

$$E_q(z) = \frac{1}{64} \left(\frac{\Delta}{2}\right)^2 \sum_i \sin^2\left(\frac{y_i}{\Delta} \pi\right) \quad (3)$$

で近似した。この関数は量子化誤差を見積もったものなので、小さくなることが望ましい。

2.2 エントロピー見積もり関数の設定

「エントロピー」を見積もる関数を設定する。

各 DCT 係数は一般的に量子化処理された後、エントロピー符号化が適用される。そこで、エントロピー符号化された DCT 係数の符号量を見積もることを考える。そのためには以下の条件を満たしていることが理想である。

- 必要な符号量を解析的に見積もりたいため、量子化された各 DCT 係数 (d_i とする) に対し、エントロピー符号化された符号量がある特定の1次微分可能な関数 $f(d_i)$ で表現できる。
- Δ が大きくなるほど量子化された各係数が0になる確率が高くなる。すなわち、0のエントロピーが減少する。

そこで、 d_i を2進表現してその桁数および2進表現そのものを符号化することにより、エントロピーを類推する。この場合の符号量は(符号ビットも含めて) $2[\log_2(d_i + 1/2) + 1] + 1[\text{bit}]$ で表せる。このステップ関数の各ステップの中央値を取る関数を使用することでエントロピーを近似する。また $d_i=0$ 近傍では上記の条件を満たすためにステップサイズ Δ に反比例した値を取ることとする。以上を考慮に入れ、エントロピー見積もり関数を以下のように定義することとする。

$$E_e(z) = \frac{1}{64} \sum_i (f(d_i)) \quad (4)$$

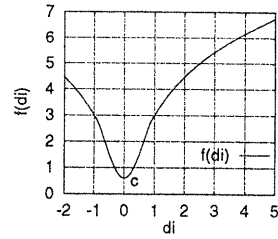


図 2: 評価関数 $f(d_i)$

$$f(d_i) = \begin{cases} 2 \log_2 \left(\frac{2|d_i|+1}{3} \right) + 3 & (|d_i| \geq 1) \\ c_4 d_i^4 + c_2 d_i^2 + c & (|d_i| < 1) \end{cases} \quad (5)$$

$$c_4 = \frac{2}{3 \log 2} - 3 + \frac{1}{2}c$$

$$c_2 = 6 - \frac{2}{3 \log 2} - \frac{3}{2}c$$

$$c = \frac{4}{\Delta}$$

$f(d_i)$ を図示すると図 2 のようになる。

2.3 評価式最適化手法

本方式においては、先に述べた量子化誤差見積もり関数 $E_q()$ 、およびエントロピー見積もり関数 $E_e()$ を同時に小さくすることが目的となる。そこで、ビットレートおよび歪みの関係から最適な padding を施すために各評価式を $E_q()$ と $E_e()$ の線形結合として表す。

$$E(z) = E_e(z) + \beta E_q(z) \quad (6)$$

なお、 β は単位 $[\text{dB} \cdot (\text{bit})^{-1}]$ を持ったビットレートに関わる定数である。ここで、評価関数はともに非線形であり、制約のない最小化問題

$$\operatorname{argmin}_{z \in Z^N} E(z) \quad (Z \text{ は有限の実数空間}) \quad (7)$$

となる(いわゆる「非線形計画問題」である)。計算量、および収束性などの観点から、このような問題の解法として最も有効であると言われている BFGS(Broydon-Fletcher-Goldfarb-Shanno)法 [2]¹ を用いて最適化を行うこととする。なお、BFGS 法を図 3 に簡単に示す。

¹Newton 法における Hesse 行列の逆行列がある行列を用いて近似し、高速化した方法(準 Newton 法)の一つ。

3 単位ブロックにおけるシミュレーション実験

前章の提案手法の評価をビットレート-歪み特性最適化の観点から行うために、計算機シミュレーションを行った。

幾つかの β に対し $\operatorname{argmin}_z E(z)$ となる z を前節の方式を利用して求めた。そして、この z を使って padding を行い、前節と同じ方法で必要ビット数、SNR を測定した。ここでは $\beta = 1$ 、および 0.01 の場合についてのみ示す(グラフ中では Proposed(β の値) と書く)。

比較のため、 X_b にすべて 0 を入れる方法 (0-P)、MRP 法 (MR-P)、MPEG-4 VM-12 における Low Pass Extrapolation Padding (LPE-P) を行い、全く同じ条件で各値を測定した。これらのグラフから、提案手法はビットレート-SNR 特性の観点から、他

手法と比較し有効であることが言える。

当然、埋め込むべき値が 1 つ (図 4(a)) の場合は変換係数にあまり差が現れないため、特性の違いはほとんど現れないが、図 4(b)(c) の $|X_b| = 16, 49$ の場合には、LPE 法と比較し大幅に効率が良くなることわかる。特にビットレート 1~2 [bit/pel] のとき、性能の向上が顕著にあらわれている。

ここで、(6) 式から β を $E_q()$ と $E_e()$ の各見積もり

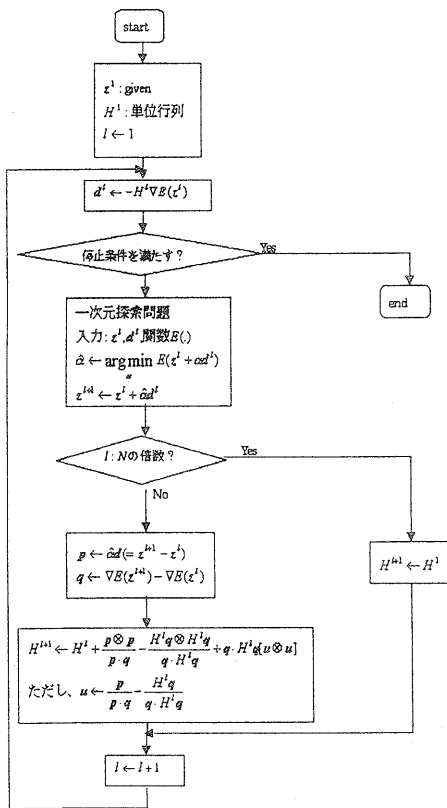
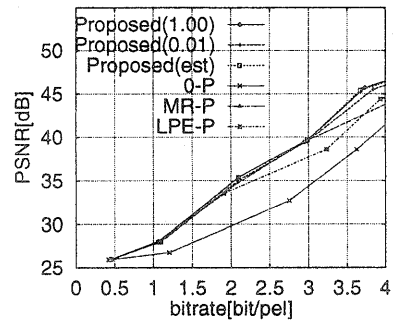
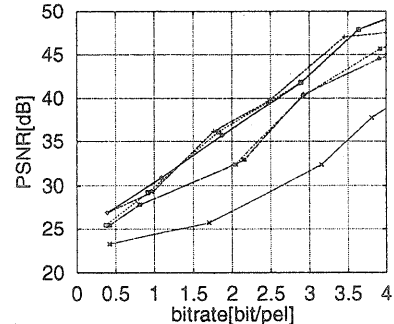


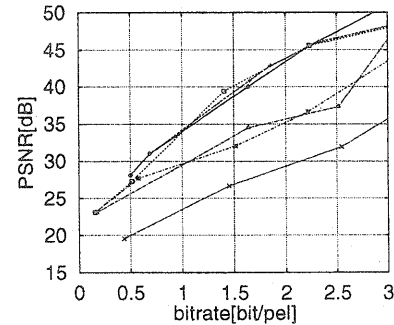
図 3: BFGS 法



(a) $|X_b| = 1$ の場合



(b) $|X_b| = 16$ の場合



(c) $|X_b| = 49$ の場合

図 4: 計算機シミュレーション結果 (8x8)

表 1: 領域分割形状と適用 DCT

形状 No.	DCT 適用形状
1,2,3,4	$N \times N$
5,8	$2N \times N$
6,7	$N \times 2N$
9,10,11,12	$2N \times 2N$ (with padding)
13	$2N \times 2N$

値の比であると考えられる。具体的には、64 個の画素値の持つ電力が DCT を適用することで、 M 個の要素に集中したと仮定する。また、量子化誤差は一様に分布すると仮定する。そして、 β を以下の式で見積もることとした。

$$\beta = \frac{(\text{エントロピーの見積もり})}{(\text{量子化誤差の見積もり})} = \frac{(f(\frac{\sqrt{64x}}{\Delta}) + Rf(\frac{\sqrt{64\sigma^2/R}}{\Delta})) / 64}{\Delta^2 / 12} \quad (8)$$

なお、 x^2 は既知物体画素値の 2 乗平均、 σ^2 は分散である。また、 R に関しては付録 A の仮定から $R = (64 - 1) \exp(-\frac{\sqrt{2}}{\sigma} \Delta)$ とおいた。また、関数 $f(\cdot)$ は式 (8) を用いた。この β を利用して padding をした場合の結果を図 4 の「Proposed(est)」に示す。 $\beta = 0.01, 1$ の場合、または従来法と比較し、式 (8) で与えられた β による padding はビットレート-歪み特性の観点から有効であることがわかる。

以下ではこの β の値を用いることとする。

4 Intra picture への適用

本章では、前章までの padding 法を用いたフレーム内符号化法を提案する。

以下、具体的に処理手法を示す。

1. $2N \times 2N$ 画素矩形領域に注目する (これをここでは「ブロック」と呼ぶ)。
2. この矩形領域を $N \times N$ 画素の 4 つの領域に分割し、図 5 上部に与えられた 13 種類の形状すべてに対して表 1 に与えられた DCT を適用する。「with padding」と書かれた矩形以外の形状のもの (No.9~12) に関しては、前節に述べた padding 法を用いて埋め込むべき値を導出する。

また、DC 予測は前のブロックから求めた DC 成分を用い、初期値 (リセット値) として TM5 [5] と同じ値を使用することとする。

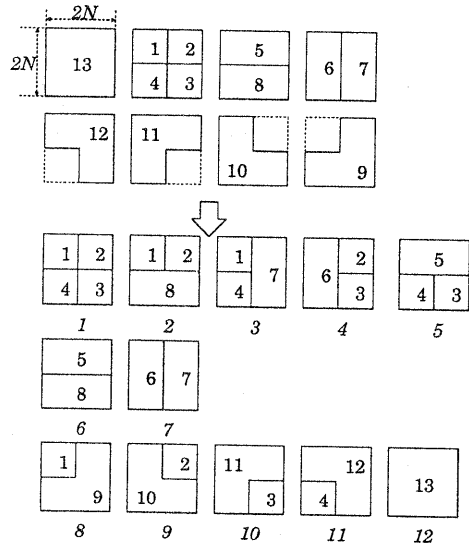


図 5: 領域分割形状とその割当

このようにして各形状に各々 DCT を適用して量子化し、各形状における必要符号量を出力する。また、復号することで量子化誤差を出力する。すなわち、各形状に対する符号量、誤差の表が出来上がる。

3. 図 5 の下部のように各領域を組合せ、必要符号量および量子化誤差を加算し、各組合せの中で (ビットレート-歪み特性の関係で) 最適な組合せを出力する。

ところで、領域分割した場合の (Intra) ブロック全体の符号量は予測誤差電力符号量 E_e 、形状情報符号量 E_s が大部分を占めるため、これらの和で表されると考えて良い。そのため、量子化誤差見積もりを E_q とした時、本来ならば E_s を何らかの形で定め、²

$$(E_e + E_s) + \beta E_q \quad (9)$$

を最小とする形状を決定すべきである。しかし、本論文では簡単のため、すべての分割形状が同じ確率で出現すると仮定し、各形状にハフマン符号を割り当てるとする。つまり、分割形状が 12 種類なの

² 画像の何らかの性質から各形状の出現確率を計算しそれに応じてハフマン符号を割り当てるのが常套手段であり、そうすることで形状情報符号量を最適にすることができるのであるが... この点は今後の課題として残しておく。

表 2: テストシーケンス

画像 フォーマット	標本化ビット数: 8bit/pel, 4:2:0, 画素数: 240 × 704/field
シーケンス	Susie(S) 女性の顔のアップ. Football(F) 各個体は大きい. Cheer Leaders(C) 観客など, 細かい個体が多い. Marching(M) 非常に細かい個体が多い.

で, 1 画素あたり

$$E_s = \frac{\log_2 12}{(2N)^2} \simeq \frac{0.896}{N^2} [\text{bit/pel}] \quad (10)$$

を加える.

5 計算機シミュレーション

前章の方式を確認すべく, 計算機シミュレーションを行った.

前章における N を 4 と置いた. テスト画像として, 今後動画の Intra Macroblock の符号化に適用していくことを考慮に入れ, 表 2 にあげた動画のある 1field を Intra field と仮定し, これをシミュレーションに用いた.

まず始めに, 量子化ステップサイズ Δ は固定とし, エントロピー符号化には JPEG で使われているハフマン符号を用いた. ここで $\Delta = 20, 40, 80$ と固定した場合のビットレート-SNR 特性を図 6 に示す. 提案手法の特性を Proposed, 従来法 (すなわち通常通り 8 画素 × 8 画素の DCT を適用した場合) の特性を Conventional とする. これを見ると判るとおり, 従来法と比較して符号化効率が良くなっていることがわかる. 特に, 絵によっては 1bit/pel 以下の低ビットレートにおいても大幅に SNR が改善されていることがわかる. ただ, 本稿では形状情報符号量を差別化していないため, 極低ビットレートにおいては形状情報符号量が強く働いてしまい, 性能が逆に悪化してしまう場面も見られた. しかしこれは何度も言うように, 形状情報符号量を差別化すれば問題ない話である. 逆にビットレートがそれ以上の場合には差別化することでさらなる改善が見込めよう.

図 7 に計算機シミュレーションにより与えられた領域分割結果を示す. 白線が領域分割形状を示して

いる. 図 7 から, 主観的に平坦な領域に対して, 量子化ステップサイズ (Δ) が小さくなるほど細かく分割されていることがわかる. 逆に Δ が大きいときは, エッジの存在する部分で領域分割されていることがわかる.

さらにどの領域分割がどの程度使われたのかを表 3 に示す. これを見ると主観的に細かいテキストチャに関しては細かく分割が行われていることがわかる. また, 提案した padding 手法を使い符号化された部分 (No.8~11) は全体の約 2 割を占めていることがわかる. 例えば (C) $\Delta = 40$ において形状 No.8~11 が選択された領域 (全 420 ブロック) におけるビットレートおよびそのときの PSNR はそれぞれ 0.83bit/pel, 31.5dB (従来の 8 × 8 画素 DCT によると 0.88bit/pel, 30.4dB) となり, 2 で提案した padding 法が有効であることがわかる.

最後に, (a) 提案手法により符号化された I Picture, および (b) TM5 により符号化された I Picture の復号画像を図 8 に示す. これらを比較すると, やはり通常の DCT で問題になっているモスキートノイズが本手法によりかなり改善されていることがわかる. 特に平坦部の近くにあるエッジによるモス

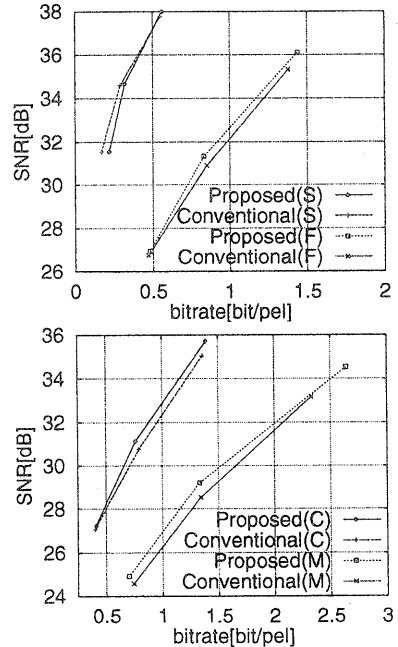


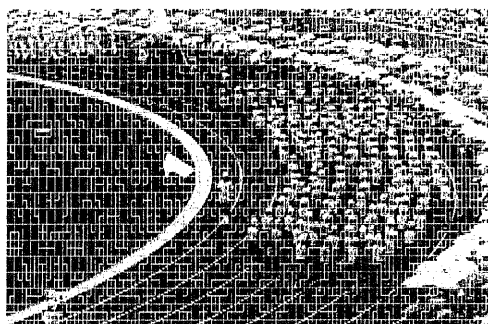
図 6: イントラピクチャのビットレート-SNR 特性 (Δ 固定)

キートノイズが本方式により低減されている。

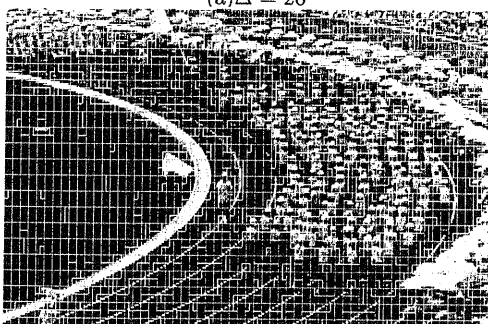
表 3: 領域分割分布

$((240/8) \times (704/8))=2640$ 領域の分布

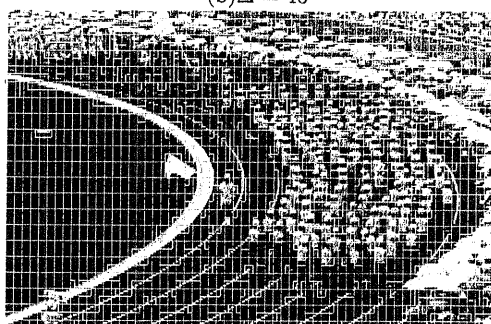
2N*2N 分離数	No. Δ	(S)	(F)	(C)		(M)
		20	40	40	80	80
4	1	2	15	7	6	15
	2	23	80	79	35	92
3	3	21	30	19	13	12
	4	30	23	19	16	22
	5	33	69	74	31	85
2	6	348	747	745	498	817
	7	135	135	152	116	45
2 with padding	8	91	137	111	94	107
	9	74	142	101	89	92
	10	83	147	107	101	107
	11	91	112	101	82	86
1	12	1709	1003	1125	1559	1160
bit rate[bit/pel]		0.56	0.76	0.83	0.48	0.70
PSNR[dB]		38.0	31.1	31.3	27.0	24.9



(a) $\Delta = 20$



(b) $\Delta = 40$

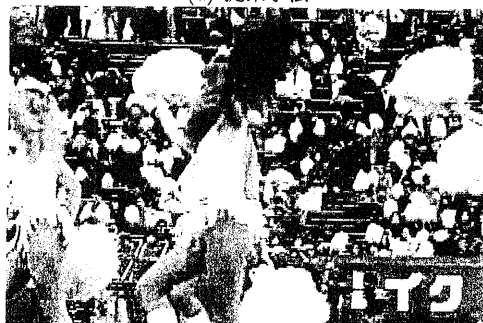


(c) $\Delta = 80$

図 7: 領域分割結果



(a) 提案手法



(b) 従来法 (8x8 DCT による)

図 8: 復号された静止画

6 おわりに

本稿では、ビットレート-歪み特性を考慮した任意形状領域符号化のための padding 手法を提案した。MPEG-4 VM で使われている LPE と比較し 1bit/pel で約 4~5dB 性能が改善できることがわかった。なお、高ビットレートではそれ以上の性能改善が見込め、0.5bit/pel ほどの低ビットレートでも同等もしくは約 1dB ほど改善できることがわかった。

また、同 padding 方式を用いた準最適フレーム内符号化方式を提案した。それにより、同一ビットレートに対し 0.5bit/pel 以上の部分で平均 0.5~0.6dB ほど性能が向上することが確認できた。

今後の課題として、既に形状が与えられた任意領域の動き推定への本 padding 手法の適用方法の検討、および、動き補償領域分割符号化への本方式の適用方法の検討などがあげられる。

なお、本研究は通信・放送機構からの委託研究「ISDB 技術に関する研究開発」に基づき行われたものである。

参考文献

- [1] R.C.Reininger and J.D.Gibson, "Distributions of the Two-Dimensional DCT Coefficients for Images," IEEE Trans. Commun.,vol.COM-31,pp.835-839.,Jun.1982.
- [2] 坂和, "非線形システムの最適化", 森北出版,1986.
- [3] 例えば,K.R.Rao and P.Yip, "Discrete Cosine Transform Algorithms, Advantages, Applications," ACADEMIC PRESS INC.,1990.
- [4] J.Kimura and M.Takizawa, "Image compression using truncated block DCT," PCS90,pp.13.1-1-2,Mar.1990.
- [5] ISO/IEC JTC1/SC29/WG11 N0400, "Test Model 5," Apr.1993.
- [6] Sikora T. and Makai B., "Shape-adaptive DCT for generic encoding of video", IEEE Trans. on Circuits and Systems for Video Technology,5,1,1995.
- [7] 小林,伊東,宇都宮:"領域分割と Warping 予測に基づいた動き補償方式," 1995 信学ソ大,D-109,Sep.1995.
- [8] 鹿喰:"領域サポート DCT を用いた任意形状符号化の検討," 信学技報,IE95-109,Dec.1995.
- [9] 三坂,中屋:"DCT 基底が持つ対称性を利用したパディング方式," 1997 信学総大,D-11-34,Mar.1997.
- [10] ISO/IEC JTC1/SC29/WG11 N2552, "MPEG-4 Video Verification Model Version 12.1," Dec.1998.

A 第2章の R の見積もり

まず,AC 係数はラプラス分布に従い [1], 平均値はほぼ 0 となると仮定する。既知画素 $\{x_i\}$ をもつ 8×8 矩形領域に DCT を適用した場合の各変換係数を $\{y_i\}$, その中の DC 成分を y_{DC} とすると AC エネルギーの総和の見積もり E_{AC} は

$$\begin{aligned} E_{AC} &= \sum_i y_i^2 - y_{DC}^2 \\ &\simeq 64\bar{x}_i^2 - (\sqrt{64}\bar{x}_i)^2 \\ &= 64\{x_i^2 - \bar{x}_i^2\} \\ &= 64\sigma^2 \end{aligned} \quad (11)$$

ただし, σ^2 は既知画素の分散, $\bar{\cdot}$ は平均を取ることを表す。ここで AC 係数の平均を 0 と仮定しているので, AC 係数の分散の見積もり σ_{AC}^2 は

$$\begin{aligned} \sigma_{AC}^2 &= \frac{E_{AC}}{64-1} \\ &= \frac{64}{64-1}\sigma^2 \\ &\simeq \sigma^2 \end{aligned} \quad (12)$$

AC 係数がラプラス分布に従うとの仮定から, 区間 $(-z, z)$ に AC 係数が存在しない確率の見積もりは

$$\begin{aligned} &1 - \int_{-z}^z \frac{1}{\sqrt{2}\sigma_{AC}} \exp\left(-\frac{\sqrt{2}}{\sigma_{AC}}|y|\right) dy \\ &= \exp\left(-\frac{\sqrt{2}}{\sigma}z\right) \end{aligned} \quad (13)$$

AC 係数は $(64-1)$ 個なので, 2章ではこれに $(64-1)$ を乗じ, さらに $z = \Delta$ とした。