

信号値の疎性を考慮した動的計画法ベース最適量子化の演算量削減 Complexity reduction for dynamic programming based quantization considering sparseness of signal value

坂東幸浩[†], 高村誠之[†], 清水淳[†]

Yukihiro BANDO[†], Seishi TAKAMURA[†] and Atsushi SHIMIZU[†]

1 はじめに

量子化 [1] の目的は、所与の評価尺度に対して、同評価尺度を最小化するための量子化インデックスを求めることにある。評価尺度として、量子化により発生する誤差（量子化誤差）が用いられる場合、量子化器の最適設計は、量子化誤差を最小化する量子化インデックスを求める最小化問題に帰着される。量子化誤差の代表的な例は、量子化前後の差分信号の二乗和である。

量子化器の設計を目的として、上述の最小化問題を求解するアプローチは、解析的に最適解を求める手法と計算機上での離散数値演算に基づき最適解を求める手法に大別される。量子化対象データの確率密度関数がパラメトリックに表現できる場合は、解析的に最適解を求める手法が適用可能な場合もある。例えば、確率密度関数が、一様分布、ガウス分布、ラプラス分布に従う場合が、それにあたる。しかし、一般的に、量子化対象データに対して、そうしたパラメトリックな表現は成立しない。

そこで、多くの場合、離散数値演算に基づく最適化手法が用いられる。代表的な手法が、Lloyd-Max 量子化法（以後、LM 量子化法と略記）[2] [3] である。しかし、LM 量子化法には、以下の二つの問題点が指摘されている。まず、第一に、LM 量子化法により得られる解には、最適性が保証されない。これは、LM 量子化法が、最適量子化の必要条件のみを考慮して設計された手法であることに起因する。LM 量子化法が最適解を与えるのは、入力データの確率密度関数の対数関数が凸性を満たす場合に限定される。なお、確率密度関数が、一様分布、ガウス分布、ラプラス分布に従う場合は、この条件に該当し、LM 量子化法を用いて最適解を得ることが可能である。一方、上記の条件を満たさない入力データを対象とした場合（大半のデータは満たさない）、繰返し処理における初期データに依存して、局所解に陥る可能性を含んでいる。第二に、収束条件を満たすまでに必要な演算回数が入力データと繰返し処理における初期データに依存するため、演算量を見積もることができないという問題もある。

一方、局所解に陥ることなく最適解を求めるアプローチとして、動的計画法を用いた適応量子化法（以後、DP 量子

化法と呼ぶ）が提案されている。Bruce[4] は、量子化器の最適化において、動的計画法における最適性の原理が適用可能なことを見出し、量子化器の最適化に必要な計算量を指数オーダーから多項式オーダーへ低減させた。Sharma[5] は、[4] に対する演算量低減策として、量子化誤差を定める歪関数が凸性制約を満たす場合に、同歪関数の最小値探索の計算量を低減するアルゴリズムを提案している。Wu[6] は、[4] に対する演算量低減策として、DP 量子化における最適解探索に matrix search 法を適用することで、同探索の計算量を低減するアルゴリズムを提案している。

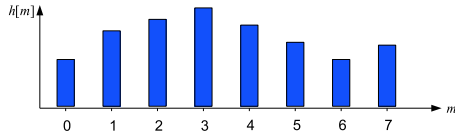
量子化器への入力が画像信号の場合、画像信号の留意すべき特性として、画素値の疎性がある。画素値の疎性とは、例えば、10 ビット信号の場合、画像内に含まれる画素値の種類が 1024 通り未満となる性質を指す。つまり、画素の振幅方向に着目すると、画像によっては、与えられたビット深度で表現可能な画素値を全て含まない場合がある。こうした画像では、画素値の頻度をヒストグラムとして表現した場合、頻度が零値になる画素値が存在する。頻度が零値になる画素値は、量子化誤差に影響を与えない。このため、頻度が零値の画素値にかかる処理を適切に省略することで、量子化誤差を最小化の際の演算量を低減できる可能性がある。例えば、画像符号化では、画素値の疎性に着目することで、符号化効率の向上を実現している [7] [8] [9]。しかし、従来の DP 量子化法では、こうした入力信号の疎性を考慮した設計がなされておらず、計算量の低減に改善の余地を残す。そこで、本稿では、画素値の疎性を考慮することで、量子化誤差の最小化を保持しつつ、演算量を低減する量子化方法を提案する。

本稿の構成は、以下の通りである。まず、2 では、量子化器の最適化問題を定式化し、本稿において解くべき問題を明確にする。次に、3 では、提案法のベースとなる DP 量子化の概要を示す。続いて、5 では、提案法であるスパース DP 量子化のアルゴリズムを示す。6 では、提案法の性能評価を行うための実験結果を示し、最後に、7 において、本稿をまとめる。

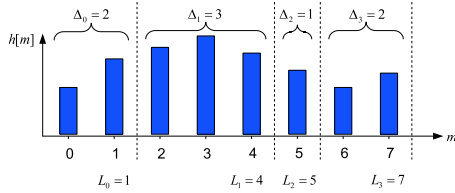
2 量子化器設計の定式化

本節では、離散信号の離散化レベルである階調数を K 階調から M 階調 ($M < K$) に量子化する場合を対象として、問題

[†]日本電信電話株式会社 NTT メディアインテリジェンス研究所



(a) 8 階調信号のヒストグラムの例



(b) 上図 (a) のヒストグラムを 4 階調に量子化した例

図 1: 量子化を表現するパラメータの例

を定式化する。なお、本定式化は、信号値 k ($k = 0, \dots, K-1$) の頻度 $h[k]$ を格納したヒストグラムを用いて行う。このとき、量子化は、2 種類のパラメータ (Δ_m と L_m) を用いて表される。 Δ_m は量子化後のヒストグラムの第 m 区間の区間幅を表す。 L_m は、同ヒストグラムにおける第 m 区間の上端であり、次式で定義されるものとする。

$$\begin{cases} L_m = \sum_{j=0}^m \Delta_j - 1 & (m = 0, \dots, M-2) \\ L_{M-1} = K - 1 \end{cases} \quad (1)$$

ヒストグラムの第 m 区間 $[L_m - (\Delta_m - 1), L_m]$ を第 m ピンと呼ぶ。なお、量子化後の各ピンが、少なくとも 1 つ以上の要素を含む必要があることから、 L_m ($0 \leq m \leq M-2$) は以下の範囲に制限される。

$$m \leq L_m \leq K - (M - m) \quad (2)$$

図 1 に $8 (= K)$ つのピンからなるヒストグラムを $4 (= M)$ つのピンからなるヒストグラムへ量子化した場合の例を示す。同図 (b) より、量子化前のピンが、各々、 $2 (= \Delta_0)$ 個、 $3 (= \Delta_1)$ 個、 $1 (= \Delta_2)$ 個、 $2 (= \Delta_3)$ 個ずつ統合され、量子化後のピンを形成していることが分る。また、量子化後のピンの上端は、各々、 $L_0 = \Delta_0 - 1 = 1$ 、 $L_1 = \Delta_0 + \Delta_1 - 1 = 4$ 、 $L_2 = \Delta_0 + \Delta_1 + \Delta_2 - 1 = 5$ 、 $L_3 = \Delta_0 + \Delta_1 + \Delta_2 + \Delta_3 - 1 = 7$ となることも確認できる。

量子化の目的は量子化誤差を最小化することである。ここで、量子化誤差としては、ヒストグラムの区間 $[L_m - (\Delta_m - 1), L_m]$ を代表値 $c(L_m - (\Delta_m - 1), L_m)$ で近似した場合の二乗誤差 $e(L_m - (\Delta_m - 1), L_m)$ を用いるものとし、次式のように定義する。

$$\begin{aligned} e(L_m - (\Delta_m - 1), L_m) \\ = \sum_{k=L_m-\Delta_m+1}^{L_m} \{k - \hat{c}(L_m - (\Delta_m - 1), L_m)\}^2 h[k] \end{aligned} \quad (3)$$

ここで、 $\hat{c}(L_m - (\Delta_m - 1), L_m)$ は、ヒストグラムの区間 $[L_m - (\Delta_m - 1), L_m]$ における重心位置 $c(L_m - (\Delta_m - 1), L_m)$ を次式により求め、さらに四捨五入して整数化した値である。

$$c(L_m - (\Delta_m - 1), L_m) = \frac{\sum_{k=L_m-(\Delta_m-1)}^{L_m} kh[k]}{\sum_{k=L_m-(\Delta_m-1)}^{L_m} h[k]} \quad (4)$$

量子化において求めるべきパラメータは、二乗誤差和を最小化する次式の M 個のパラメータである。

$$\begin{aligned} & (\Delta_0^*, \dots, \Delta_{M-1}^*) \\ & = \arg \min_{\Delta_0, \dots, \Delta_{M-1}} \left\{ \sum_{m=0}^{M-1} e(L_m - (\Delta_m - 1), L_m) \right\} \end{aligned} \quad (5)$$

3 DP 量子化

従来研究に対する提案法 (5 節にて後述) の差分を明確化し、提案法に対する理解を促進するために、ベースとなる DP 量子化の概要を示す。式 (5) における M 個のパラメータ ($\Delta_0, \dots, \Delta_{M-1}$) の取り得る組み合わせは、 M の増加と共に指数関数的に増加するため、最適な組み合わせ ($\Delta_0^*, \dots, \Delta_{M-1}^*$) を総当りで探索するのは、計算量的に困難である。そこで、第 m ピンの量子化誤差 $e(L_m - (\Delta_m - 1), L_m)$ が同ピンの上端 L_m と区間幅 Δ_m に依存することに着目し、以下のように、動的計画法を用いて最適解を求めるアプローチ (以下、DP 量子化と呼ぶ) が導入された [4] [5] [6]。

DP 量子化では、量子化誤差に対する漸化関係に着目する。ヒストグラムの区間 $[0, L_m]$ を $m+1$ 個のピンで構成した際の量子化誤差の総和 $\sum_{i=0}^m e(L_i - (\Delta_i - 1), L_i)$ の最小値を参照テーブル $S_m[L_m]$ に格納するものとする。つまり、最適な $\Delta_m, \dots, \Delta_0$ を用いた場合の $\sum_{i=0}^m e(L_i - (\Delta_i - 1), L_i)$ に対する最小値である。ここで、 $e(L_m - (\Delta_m - 1), L_m)$ が第 m ピンの上端 L_m と区間幅 Δ_m に依存することから、 $S_m[L_m]$ は $S_{m-1}[L_m - \Delta_m]$ を用いて、次式のように表わせる。

$$\begin{aligned} S_m[L_m] &= \min_{\Delta_m} [S_{m-1}[L_m - \Delta_m] \\ &\quad + e(L_m - (\Delta_m - 1), L_m)] \end{aligned} \quad (6)$$

なお、 $m = 1, \dots, M-1$ である。また、 $L_m = m, \dots, K - (M - m)$ である。漸化式 (6) を用いることで、 $S_m[L_m]$ の算出は、 $\Delta_m = 1, \dots, L_m - m + 1$ からの最適なパラメータ選択問題に帰着させられる。 Δ_m の取り得る範囲については、付録 A を参照されたい。ここで、 $S_m[L_m]$ に格納した値は $S_{m+1}[L_{m+1}]$ の算出において用いるものとする。

式 (5) の最小化問題は、次式のように表せる。

$$\begin{aligned} \min_{\Delta_{M-1}} [& S_{M-2}[L_{M-1} - \Delta_{M-1}] \\ & + e(L_{M-1} - (\Delta_{M-1} - 1), L_{M-1})] \end{aligned} \quad (7)$$

上式の最小化問題は、 $\frac{(K-M)(1+K-M)(M-1)}{2}$ 個の候補の中から最適解 ($\Delta_0^*, \dots, \Delta_{M-1}^*$) を求める最適化問題に帰着される。

この結果、提案法は多項式時間での最適解の求解が可能となる。

4 DP量子化に対する最適経路探索としての解釈

DP量子化の最適化のメカニズムを理解するために、トレリス遷移図を用いた解釈を与える。これは、後述する提案アルゴリズムの見通しを良くするためである。図1(b)に示した量子化の例を図2に示すトレリス遷移図を用いて表現する。同図横軸 k は入力信号値、同図縦軸 m は量子化後の値を表している。また、同図における (k, m) の位置のノードには、ヒストグラムの区間 $[0, k]$ を $m+1$ 階調に量子化した場合の量子化誤差和の最小値を格納する。例えば、同図の $(4, 1)$ の位置にするノードには、ヒストグラムの第0から第4要素までの5要素を2階調に量子化した場合の量子化誤差和の最小値が格納される。なお、同図の最左下のノードは、トレリス遷移図における始点として導入したダミーノードであり、物理的な意味は有しない。図1(b)の量子化は、図2の太線で示す経路に対応する。このようにトレリス遷移図で考えると、最適な量子化設計は、トレリス遷移図上での最適経路探索問題となる。

また、ノード間のパスには、対応するヒストグラムの区間を量子化した際の量子化誤差を付与するものとする。例えば、ノード $(k, m)=(1, 0)$ から $(k, m)=(4, 1)$ へのパスには、2つ目の量子化ピンとしてヒストグラムの区間 $[1, 4]$ を量子化した場合の量子化誤差が付与される。

動的計画法を用いることで計算量が低減できる理由は、トレリス遷移図を用いると、以下のように解釈できる。例えば、図3の黒色ノードに至る最適経路は、始点から同ノードに至る全経路の中から量子化誤差和を最小化する最適経路を探索する必要はない。一段下の灰色ノードから黒色ノードに至る経路に限定して、探索すれば十分である。灰色ノードに至るまでの量子化誤差和の最小値は、灰色ノードに格納されている。このため、灰色ノードに格納された量子化誤差和に灰色ノードから黒色ノードに至るパスに付与された量子化誤差を加えれば、同灰色ノードを經由して黒色ノードに至る場合の量子化誤差和の最小値が求まる。そこで、4つの灰色ノードに対して黒色ノードに至る場合の量子化誤差和を求め、同量子化誤差和の最小値を黒色ノードに格納すればよい。

5 スパースDP量子化

5.1 着眼点

ヒストグラムの非有意要素(頻度が零値となる要素)に着目したDP量子化の演算量削減手法を導入する。画素値

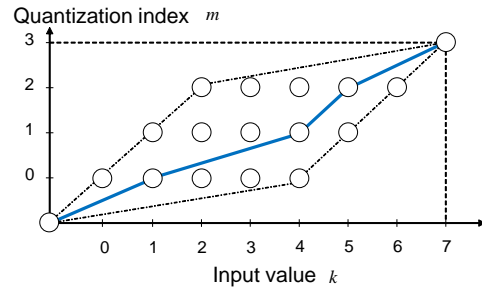


図2: 図1(b)の量子化に対応するトレリス図における経路

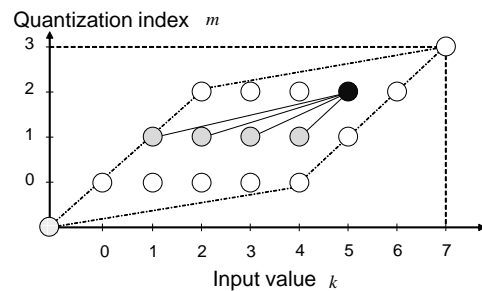


図3: 黒色ノードに対する最適経路探索

$L_m + 1$ における頻度 $h[L_m + 1]$ が零値である場合、即ち、 $h[L_m + 1] = 0$ の場合、ヒストグラムの区間 $[L_m - (\Delta_m - 1), L_m + 1]$ に対する量子化誤差は、ヒストグラムの区間 $[L_m - (\Delta_m - 1), L_m]$ に対する量子化誤差と等しい。 $h[L_m + 1] = 0$ であることから、量子化の対象となる区間に $h[L_m + 1]$ を加えたとしても、量子化による近似には影響を与えないためである。従って、 $h[L_m + 1] = 0$ の場合、前述の式(6)に基づく $S_m[L_m + 1]$ の算出は省略することが可能となる。つまり、量子化誤差の最小化において考慮すべきは、ヒストグラムの有意要素(頻度が非零値となる要素)のみで良いことが分かる。画素値の疎性に関する仮説を検証するために、全要素数に対する非有意要素数の比として、疎性を次式のように定義し、

$$\text{疎性度} = \frac{\text{非有意要素数}}{\text{全要素数}} \quad (8)$$

標準画像(仕様の詳細は6節にて後述)に対して、疎性を計測したところ、表1に示す結果となり、画素値に関する一定の疎性が確認できた。

そこで、ヒストグラム $h[k]$ ($k = 0, \dots, K-1$)の要素を指定するインデックス k を当該ヒストグラムの要素インデックスと呼び、あわせて、ヒストグラム $h[k]$ の有意要素(\tilde{K} 個)を指定するインデックス \tilde{k} ($\tilde{k} = 0, \dots, \tilde{K}-1$)を当該ヒストグラムの有意要素インデックスと呼び、有意要素インデックスを用いて量子化アルゴリズムを構成する。この量子化アルゴリズムを記述する為に、以下の参照テーブルを

表 1: 各量子化対象画像の疎性度 (表中の Sparseness の列は式 (8) で定義された値)

Quantized signal	Sparseness [%]
Image1	7.5
Image2	8.2
Image3	11.2
Image4	19.6

用意する。\$F[\tilde{k}]\$ は、第 \$\tilde{k}\$ 番目の有意要素に対応する要素インデックスを格納する。\$\Psi_u[m]\$ (\$m = 0, \dots, M-1\$) は第 \$m\$ ビンの上端となりうる有意要素インデックスの最大値を格納する。\$\Psi_l[m]\$ (\$m = 0, \dots, M-1\$) は第 \$m\$ ビンの上端となりうる有意要素インデックスの最小値を格納する。\$\Psi_u[m]\$, \$\Psi_l[m]\$ (\$m = 0, \dots, M-1\$) は各々、次のように設定される。

$$\Psi_u[m] = \psi_u[m-M+K] - Z[\psi_u[m-M+K]] \quad (m = 0, \dots, M-1) \quad (9)$$

$$\Psi_l[m] = \begin{cases} \psi_l[m] - Z[\psi_l[m]] & (m = 0, \dots, M-2) \\ K-1 - Z[K-1] & (m = M-1) \end{cases} \quad (10)$$

ここで、\$\psi_u[m-M+K]\$ は、有意要素の要素インデックスのうち、\$m-M+K\$ 以下の範囲における最大値であり、\$\psi_l[m]\$ は、有意要素の要素インデックスのうち、\$m\$ 以上の範囲における最小値である。\$Z[k]\$ (\$k = 0, \dots, K-1\$) は、ヒストグラムの区間 \$[0, k]\$ に含まれる非有意要素の個数を格納する。

\$M = 8, K = 4\$ の量子化の場合を例にとり、同量子化に対応する図 4 のトレリス図を用いて、\$\Psi_u[m]\$, \$\Psi_l[m]\$ の例を示す。ここでは、\$k = 0, 2, 6\$ の要素が非有意要素であるとする。同図において灰色ノードは、非有意要素に対応するノードであることを示す。また、白ノード内の値は、同ノードの有意要素インデックスを表す。量子化対象の信号に非有意要素が含まれない場合、各ビンの上端は、緑色破線上のノード (第 \$m\$ ビンの場合、要素インデックス \$m\$ のノード) から青色破線上のノード (第 \$m\$ ビンの場合、要素インデックス \$m-M+K\$ のノード) の範囲に設定される。一方、非有意要素が含まれる場合は、非有意要素をビンの上端としないものとして、ビンの上端の範囲を設定する。例えば、同図の第 1 ビンの場合、同ビン上端の最小値は緑色破線上のノードであり、その有意要素インデックスは \$\Psi_l[1] = 0\$ となる。また、同ビン上端の最大値は青色破線上のノードであり、その有意要素インデックスは \$\Psi_u[1] = 3\$ となる。また、第 2 ビンの場合、緑色破線上のノードが非有意要素のため、同ビン上端の最小値は緑色破線上のノードより右側に存在するノードの中で、最も左側の白色ノードとなる。このため、対応する有意要素インデックスは \$\Psi_l[2] = 1\$ となる。また、青色破線上のノードも非有意要素のため、同ビン上端の最大値は青色破線上のノードより左側に存在す

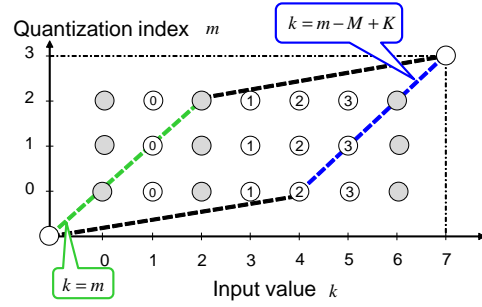


図 4: ビンの上端の範囲

るノードの中で、最も右側の白色ノードとなる。このため、対応する有意要素インデックスは \$\Psi_u[2] = 3\$ となる。

5.2 最適解の求解

提案するスパース DP 量子化は、最適性を保持した上で DP 量子化の演算量を低減するため、非有意要素に対する不要な演算の省略を特徴とする。以下、スパース DP 量子化のアルゴリズムを示す。

\$\tilde{L}_m\$ 番目の有意要素を上端とするヒストグラムの区間 \$[0, F[\tilde{L}_m]]\$ を \$m+1\$ 分割した場合を考える。同区間の分割により得られる部分区間 \$[F[\tilde{L}_i - (\tilde{\Delta}_i - 1)], F[\tilde{L}_i]]\$ (ここで、\$i = 0, \dots, m\$) を第 \$i\$ ビンとする。第 \$i\$ ビンに対して、重心で近似した場合の量子化誤差を算出し、同量子化誤差 \$e(F[\tilde{L}_i - (\tilde{\Delta}_i - 1)], F[\tilde{L}_i])\$ を参照テーブル \$E[\tilde{L}_i - (\tilde{\Delta}_i - 1), \tilde{L}_i]\$ に格納するものとする。ここで、\$\tilde{L}_i\$ は第 \$i\$ ビンの上端にあたる有意要素インデックスであり、\$\tilde{\Delta}_i\$ は第 \$i\$ ビン内の有意要素数である。さらに、量子化誤差和 \$\sum_{i=0}^m E[\tilde{L}_i - (\tilde{\Delta}_i - 1), \tilde{L}_i]\$ の最小値を格納する参照テーブルを \$\tilde{S}_m[\tilde{L}_m]\$ として定義する。つまり、最適な \$\tilde{\Delta}_m, \dots, \tilde{\Delta}_0\$ を用いた場合の \$\sum_{i=0}^m E[\tilde{L}_i - (\tilde{\Delta}_i - 1), \tilde{L}_i]\$ に対する最小値である。なお、\$\tilde{S}_m[\tilde{L}_m]\$ は \$S_m[F[\tilde{L}_m]]\$ と等しい。

\$E[\tilde{L}_m - (\tilde{\Delta}_m - 1), \tilde{L}_m]\$ が、第 \$m\$ ビンの上端の有意要素インデックス \$\tilde{L}_m\$ と同ビン内の有意要素数 \$\tilde{\Delta}_m\$ に依存することに着目すると、\$\tilde{S}_m[\tilde{L}_m]\$ に格納する値は \$\tilde{S}_{m-1}[\tilde{L}_m - \tilde{\Delta}_m]\$ を用いて、次式のように表わせる。

$$\tilde{S}_m[\tilde{L}_m] = \min_{\tilde{\Delta}_m} [\tilde{S}_{m-1}[\tilde{L}_m - \tilde{\Delta}_m] + E[\tilde{L}_m - (\tilde{\Delta}_m - 1), \tilde{L}_m]] \quad (11)$$

ここで、\$m = 1, \dots, M-1\$ である。漸化式 (11) を用いることで、\$\tilde{S}_m[\tilde{L}_m]\$ の計算は、\$\tilde{\Delta}_m = 1, \dots, \tilde{L}_m - \Psi_l[m-1]\$ から最適値を選択する問題に帰着する。\$\tilde{\Delta}_m\$ の範囲については、付録 B を参照されたい。なお、第 \$m\$ ビンにおいて取り得る有意要素インデックスの最大値および最小値を各々、\$\Psi_u[m]\$ および \$\Psi_l[m]\$ としたことから、\$\tilde{L}_m\$ の取り得る値は、\$\tilde{L}_m = \Psi_l[m], \dots, \Psi_u[m]\$ となる。\$\tilde{S}_m[\tilde{L}_m]\$ に格納した値は \$\tilde{S}_{m+1}[\tilde{L}_{m+1}]\$ の算出時に参照するものとする。式 (11)

1. Load histogram $h[k]$ ($k = 1, \dots, K - 1$) of the input signal
2. Load the number of quantization bin M
3. Generate look-up-tables $Z[k]$ ($k = 0, \dots, K - 1$), $\Psi_l[m]$, $\Psi_u[m]$ ($m = 0, \dots, M - 1$) and $F[\tilde{k}]$ ($\tilde{k} = 0, \dots, K - Z[K - 1] - 1$)
4. Generate look-up-table $E[\tilde{i}_s, \tilde{i}_e]$ ($\tilde{i}_s = 0, \dots, K - Z[K - 1] - 2$, $\tilde{i}_e = 1, \dots, K - Z[K - 1] - 1$) for quantization error of each interval in histogram $h[k]$
5. for $j = 0, \dots, K - M - 1 - Z[K - M - 1]$
6. $S_0[j] \leftarrow \tilde{e}[0, j]$
7. for $m = 1, \dots, M - 1$
8. for $\tilde{L}_m = \Psi_l[m], \dots, \Psi_u[m]$
9. $\tilde{S}_m[\tilde{L}_m] \leftarrow \min_{\tilde{\Delta}_m=1, \dots, \tilde{L}_m - \Psi_l[m-1]} [\tilde{S}_{m-1}[\tilde{L}_m - \tilde{\Delta}_m] + E[\tilde{L}_m - (\tilde{\Delta}_m - 1), \tilde{L}_m]]$
10. $\tilde{\Delta}_m^{(\tilde{L}_m)} \leftarrow \arg \min_{\tilde{\Delta}_m=1, \dots, \tilde{L}_m - \Psi_l[m-1]} [\tilde{S}_{m-1}[\tilde{L}_m - \tilde{\Delta}_m] + E[\tilde{L}_m - (\tilde{\Delta}_m - 1), \tilde{L}_m]]$
11. $T_{m-1}[\tilde{L}_m] \leftarrow \tilde{L}_m - \tilde{\Delta}_m^{(\tilde{L}_m)}$
12. $\tilde{L}_{M-1}^* \leftarrow K - Z[K - 1] - 1$
13. for $m = M - 1, \dots, 1$
14. $\tilde{L}_{m-1}^* \leftarrow T_{m-1}[\tilde{L}_m^*]$
15. $\Delta_m^* \leftarrow F[\tilde{L}_m^*] - F[\tilde{L}_{m-1}^*]$
16. $\Delta_0^* \leftarrow F[\tilde{L}_0^*] + 1$

図 5: スパース DP 量子化アルゴリズム

の右辺を最小化する $\tilde{\Delta}_m$ を $\tilde{\Delta}_m^{(\tilde{L}_m)}$ とし、第 m ビンの上端を \tilde{L}_m とした場合、第 $m - 1$ ビンの上端の最適値として、 $T_{m-1}[\tilde{L}_m] = \tilde{L}_m - \tilde{\Delta}_m^{(\tilde{L}_m)}$ を各 $\tilde{L}_m (= \Psi_l[m], \dots, \Psi_u[m])$ に対して格納しておくものとする。なお、 $m = 0$ の場合、 $\tilde{S}_0[\tilde{L}_0]$ はヒストグラムの区間 $[0, F[\tilde{L}_0]]$ を重心で近似した場合の量子化誤差であることから、

$$\tilde{S}_0[\tilde{L}_0] = E[0, F[\tilde{L}_0]]$$

となる。以上の処理を擬似コードを用いて表すと、図 5 における 4. ~ 11. となる。同図 1. ~ 3. は、5.1 で示した参照テーブルの生成である。

量子化後の各ピンを規定する $(\Delta_0^*, \dots, \Delta_{M-1}^*)$ は、以下に示すバックトラック過程を通して求める。式 (5) の最小化問題は、次式のように表せる。

$$\min_{\tilde{\Delta}_{M-1}} [\tilde{S}_{M-2}[\tilde{L}_{M-1} - \tilde{\Delta}_{M-1}] + E[\tilde{L}_{M-1} - (\tilde{\Delta}_{M-1} - 1), \tilde{L}_{M-1}]] \quad (12)$$

上式の最小化を実現する $\tilde{\Delta}_{M-1}$ を

$$\tilde{\Delta}_{M-1}^* = \arg \min_{\tilde{\Delta}_{M-1}} [\tilde{S}_{M-2}[\tilde{L}_{M-1} - \tilde{\Delta}_{M-1}] + E[\tilde{L}_{M-1} - (\tilde{\Delta}_{M-1} - 1), \tilde{L}_{M-1}]]$$

とおく。このとき、 \tilde{L}_{M-1} の取り得る値は $K - Z[K - 1] - 1$ のみであることから、 $\tilde{L}_{M-1} = K - Z[K - 1] - 1$ となる。 \tilde{L}_{M-1} および $\tilde{\Delta}_{M-1}^*$ を用いて、第 $M - 2$ ビンの上端の有意要素インデックスの最適値は、 $\tilde{L}_{M-2}^* = \tilde{L}_{M-1} - \tilde{\Delta}_{M-1}^* = K - Z[K - 1] - 1 - \tilde{\Delta}_{M-1}^*$ と求まる。第 $M - 3$ ビンの上端の有意要素インデックスの最適値は、 \tilde{L}_{M-2} に対する最適解として、 $T_{M-3}[\tilde{L}_{M-2}^*]$ に格納されているので、該当する値を参照し、 $\tilde{L}_{M-3}^* = T_{M-3}[\tilde{L}_{M-2}^*]$ とする。以下、同様の参照処理を $\tilde{L}_{M-4}^* = T_{M-4}[\tilde{L}_{M-3}^*], \dots, \tilde{L}_0^* = T_0[\tilde{L}_1^*]$ として繰り返す。ここで得られた各ビンの上端の有意要素インデックス $\tilde{L}_{M-1}, \tilde{L}_{M-2}, \dots, \tilde{L}_0$ を用いて、 $F[\]$ を参照し、各ビンの上端の要素インデックスが求まる。この結果、各ビンの区間幅は $\Delta_{M-1}^* = F[\tilde{L}_{M-1}] - F[\tilde{L}_{M-2}^*] = K - 1 -$



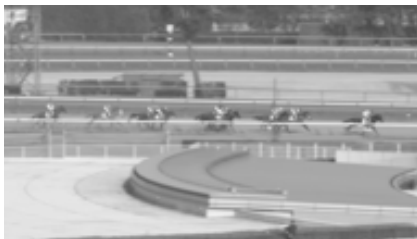
(a) Image1



(b) Image2



(b) Image3



(d) Image4

図 6: 量子化対象とした画像

$F[\tilde{L}_{M-2}^*], \Delta_{M-2}^* = F[\tilde{L}_{M-2}^*] - F[\tilde{L}_{M-3}^*], \dots, \Delta_1^* = F[\tilde{L}_1^*] - F[\tilde{L}_0^*], \Delta_0^* = F[\tilde{L}_0^*] + 1$ として求まる。以上の処理を擬似コードを用いて表すと、図 5 における 12.~16. となる。

6 実験

提案手法を評価するために画像信号を対象とした量子化を行い、処理量の観点から考察を行った。量子化の対象として、ハイビジョン・システム評価用標準動画像 第 2 版 (映像情報メディア学会/電波産業会 発行) に含まれる動画像 B シリーズの先頭フレーム (図 6) の輝度信号を用いた。対象とした動画像の仕様は、ビット深度: 10bit/channel、走査フォーマット: プログレッシブ、空間解像度: 1920 × 1080 画素である。量子化対象の動画像のビット深度が 10bit/channel

表 2: DP-Q 及び SDP-Q の処理時間 (表中の値は M 毎の平均値。reduction ratio の列の値は式 (13) で定義される値)

M	DP-Q [msec]	SDP-Q [msec]	reduction ratio [%]
128	300	230	23.5
256	324	286	11.7

であることから、原信号の階調数は $K = 1024$ となる。量子化後の階調数 M は、 $M = 128, 256$ とした。実験に用いた計算機は、CPU: Intel core i7(2.8GHz), memory: 8GB である。

スパース DP 量子化による演算量削減の効果を検証するため、文献 [4] のアルゴリズム (3 節に記載) を実装した DP 量子化 (以下の図表式では DP-Q と略記) とスパース DP 量子化 (以下の図表式では SDP-Q と略記) の処理時間を比較した。この結果を図 7 に示す。ここで、処理時間は、1000 回の処理の平均時間である。DP 量子化に対するスパース DP 量子化の演算量低減率を次式により定義する:

$$\text{演算量低減率} = \frac{\text{DP-Q の処理時間} - \text{SDP-Q の処理時間}}{\text{DP-Q の処理時間}} \quad (13)$$

全画像に対する平均的な特性を俯瞰するために、表 2 に各 M の値における平均値を示す。スパース DP 量子化は、DP 量子化に対して、平均 11.7 ~ 23.5% の処理時間の低減を実現している。

処理時間の低減率は、 M が小さい場合に高い値を示している。この理由は以下のように考察される。非有意要素の個数が一定の場合、出力ビン数である M が小さいほど、出力ビン数に対する非有意要素数の比率は大きくなる。このため、非有意要素に対する処理を省略することで演算量を削減するスパース DP 量子化では、 M が小さい場合に DP 量子化誤差に対する演算量の低減率が高くなると考えられる。

疎性度の観点からスパース DP 量子化の演算量低減を考察する。5.1 節にて表 1 に示したように、式 (8) で定義した疎性度は、“image1”, “image2”, “image3”, “image4” の順に増加している。図 7 によれば、量子化対象画像の疎性度の増加に従い、演算量低減率も向上している。

7 まとめ

本稿では、ヒストグラムの疎性に着目し、最適量子化である DP 量子化の演算量低減について検討した。提案法であるスパース DP 量子化は DP 量子化の最適性を継承しており、両量子化の量子化誤差は一致する。つまり、スパース DP 量子化導入の目的は、量子化誤差の最小化を保持した上で、DP 量子化の演算量低減を図ることにある。画像信号を用いた実験を通して、スパース DP 量子化は、従来

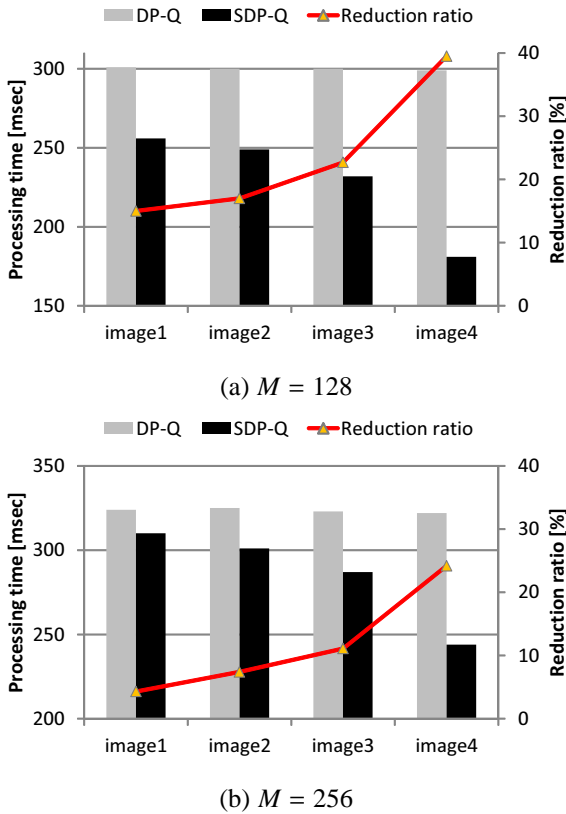


図7: DP-Q及びSDP-Qの処理時間(DP-Qの棒グラフはDP量子化の処理時間、SDP-Qの棒グラフはスパースDP量子化の処理時間、Reduction ratioは式(13)で定義される値)

のDP量子化に対して、演算量を平均11.7~23.5%低減可能となることを確認した。なお、本稿で提案した信号値の疎性を考慮したDP量子化の演算量低減手法は、DP量子化に対する既存の演算量低減手法[5][6]とも相補的な関係にあるため、これらの既存手法と組み合わせることで、さらなる演算量低減が図れるものと考えられる。

参考文献

- [1] R. M. Gray and D. L. Neuhoff. Quantization. *IEEE Trans. Information Theory*, Vol. 44, No. 6, pp. 2325–2383, Oct. 1998.
- [2] S. P. Lloyd. Least squares quantization in PCM. *IEEE Trans. Information Theory*, Vol. IT-28, pp. 129–136, Mar. 1982.
- [3] J. Max. Quantizing for minimum distortion. *IRE Trans. Information Theory*, Vol. IT-7, pp. 7–12, Mar. 1960.
- [4] J. D. Bruce. *Optimum quantizer*. PhD thesis, M.I.T., May 1964.

- [5] D. Sharma. Design of absolutely optimal quantizers for a wide class of distortion measures. *IEEE Trans. Information Theory*, Vol. 24, No. 6, pp. 693–702, Nov. 1978.
- [6] X. Wu. Optimal quantization by matrix searching. *Journal of Algorithms*, Vol. 12, No. 4, pp. 663–673, Dec. 1991.
- [7] P. Ferreira and A. J. Pinho. Why does histogram packing improve lossless compression rates? *IEEE Signal processing letters*, Vol. 9, No. 8, pp. 259–261, 2002.
- [8] M. Aguzzi and M. Albanesi. A novel approach to sparse histogram image lossless compression using JPEG 2000. *Electronic Letters on Computer Vision and Image Analysis*, Vol. 5, No. 4, pp. 24–46, 2006.
- [9] E. Nasr-Esfahani, S. Samavi, N. Karimi, and S. Shiran. Near lossless image compression by local packing of histogram. *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pp. 1197–1200, 2008.

付録A Δ_m の値域

Δ_m の取り得る範囲について考察する。 $L_{m-1} = L_m - \Delta_m$ および式(2)より、 $L_m - \Delta_m$ の範囲は、

$$m - 1 \leq L_m - \Delta_m \leq K - (M - m + 1)$$

となる。このため、 Δ_m の範囲は、与えられた L_m を用いて次式のように表せる。

$$L_m - K + (M - m + 1) \leq \Delta_m \leq L_m - m + 1 \quad (14)$$

さらに、 $\Delta_m \geq 1$ であることを考慮すると、次式を得る。

$$1 \leq \Delta_m \leq L_m - m + 1 \quad (15)$$

付録B $\tilde{\Delta}_m$ の値域

$\tilde{\Delta}_m$ の範囲は、次のようになる。 $\tilde{L}_{m-1} = \tilde{L}_m - \tilde{\Delta}_m$ であることから、 $\tilde{L}_m - \tilde{\Delta}_m$ の範囲は、 $\Psi_l[m-1] \leq \tilde{L}_m - \tilde{\Delta}_m \leq \Psi_u[m-1]$ となる。このため、 $\tilde{\Delta}_m$ の範囲は、与えられた \tilde{L}_m を用いて次式のように表せる。

$$\tilde{L}_m - \Psi_u[m-1] \leq \tilde{\Delta}_m \leq \tilde{L}_m - \Psi_l[m-1] \quad (16)$$

さらに、 $\tilde{\Delta}_m \geq 1$ であることを考慮すると、次式を得る。

$$1 \leq \tilde{\Delta}_m \leq \tilde{L}_m - \Psi_l[m-1] \quad (17)$$