

〈一般研究課題〉 画像デジタルアーカイブのための様々な再生機器解像度に対応する高能率可逆画像符号化手法に関する研究
助成研究者 中京大学 青森 久



画像デジタルアーカイブのための様々な再生機器解像度に対応する高能率可逆画像符号化手法に関する研究

青森 久
(中京大学)

Research on High Efficiency Scalable Lossless Image Coding Method for Digital Image Archiving

Hisashi Aomori
(Chukyo University)

Abstract :

This paper proposes a novel scalable lossless image coding scheme with pel-adaptive prediction using cellular neural network (CNN). The scalable image coding scheme is indispensable for modern digital archiving applications, since they are used by various mobile devices. Also, from the viewpoint of the optimal lossless coding, a pel-adaptive predictor enables high prediction performance. In this paper, edge-orientation predictors consist of space-variant CNN are used for the scalable image coding scheme having pel-adaptive prediction with no selection information. The effectiveness of proposed algorithm is validated by some computer simulations of various standard test images, and its performance is compared with that of other existing coding schemes.

1. はじめに

観光立国推進基本法の成立により、各市町村や自治体などは地域情報資源を基にした画像アーカイブを盛んに構築している。これら資源の積極的な公開やネットワークを通じた利用が容易になっているものの、デジタルアーカイブの対象は有形・無形文化遺産、メディア芸術、文化施設などに限られているという側面があった。しかし、東日本大震災の発生により、我々の身近な風景や生活風景を記録することが復興や防災計画立案上重要であると認識され、現在様々な震災関連デジタルアーカイブプロジェクトが展開されており、デジタルアーカイブ構築の重要性が増している。また、

資料の観点から、これらのアーカイブは可逆符号化方式での符号化が求められ、さらに、デジタルアーカイブ利用端末の再生解像度の多様化などにより異なる解像度の画像を再生できる符号化データが望まれている。

本研究では、様々な再生機器解像度に対応するデジタルアーカイブ構築のための高能率画像符号化手法の構築を研究の目的とする。本方式は、画像の分割処理と網膜ネットワークであるセルラーニューラルネットワーク(CNN)[1]を用いた予測処理を階層的に繰り返す構造で成り立っている。従って、復号側では符号化側と同様の処理を行うことで原画像を完全に復元可能であり、更に復号の各段階において異なる解像度の画像を再生することができる。また、輪郭の方向性に応じたCNN予測器と予測誤差を算術符号化する際の確率モデルを画素毎に適応選択する方式を採用している。ここで、CNNは人工網膜ネットワークの1種であり、画像の画素と1対1に対応したニューロンが空間的に配置され、各ニューロン同士が局所的に接続された疎結合ネットワーク構造を有する。この特徴から、CNNはフィルタリング、画像圧縮、パターン認識などの多くの画像処理アプリケーションに用いられている。

本方式の予測器設計は、階層毎に存在するCNN予測器の特徴を決定づける複数のパラメータを符号化の最終段階にて得られる予測誤差の符号量を最小化するよう各予測器を学習させることと等価である。本方式では、予測誤差を算術符号化する際の見積もり符号量を最小とするよう再生される順から予測器を反復的に最適化することで、実際の符号量が最小となる学習方法実現している。さらに、確率モデルを推定する特徴量算出の際に、階層差に逆比例した重み付けを参照する予測誤差にすることで、コンテキストモデリングにおける確率モデル推定精度の向上を図っている。しかし、深層学習の学習に多大なコストがかかることからわかるように、一般に多層ニューラルネットワークの学習には困難が伴う。本方式を実際の画像に適用することを考えると、階層数は最低でも12層以上となるため、効果的な学習方法を構築する必要がある。本研究では、最適化の際に目的関数の微分係数を必要としないメタヒューリスティックな最適化手法であるPSO(Particle Swarm Optimization)[2]を学習に用いることにより、学習の困難さに対処している。

2. CNNを用いた階層的な可逆符号化方式

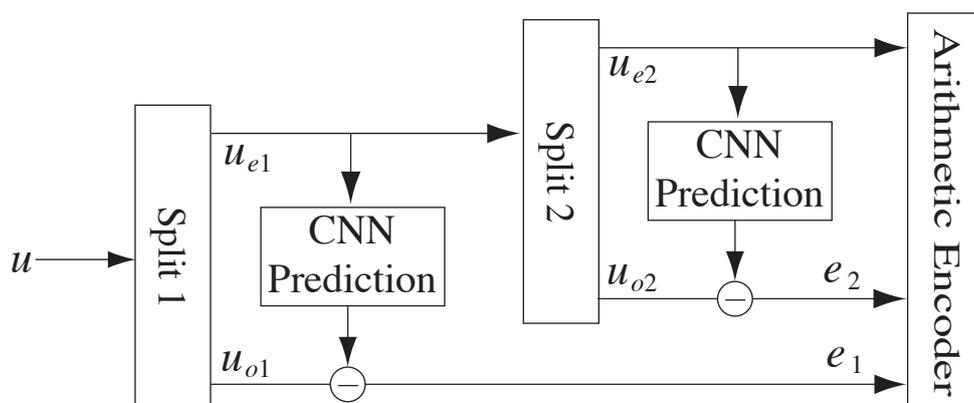


図1 CNN予測器による階層的な可逆符号化方式

図1にCNN予測器を用いた階層的な可逆符号化方式[3]を示す。本符号化方式の主な処理は画像の分割処理、CNN予測器による予測処理及び算術符号化を用いたエントロピー符号化の3ステップで構

成されている。

画像の分割処理において、入力画像 U は偶数要素 U_e と奇数要素 U_o に図2に示すquincunxサブサンプリング方式により分割される。ここで、quincunxサブサンプリングにおける2種の分割パターンをそれぞれ奇数ステージと偶数ステージと呼び、奇偶ステージをまとめて1レベルとする。このような画像分割方法に基づく階層的符号化方式を採用することにより、入力画像の多重解像度表現が実現でき、復号のレベル毎に異なる解像度の再生画像を取得可能である。

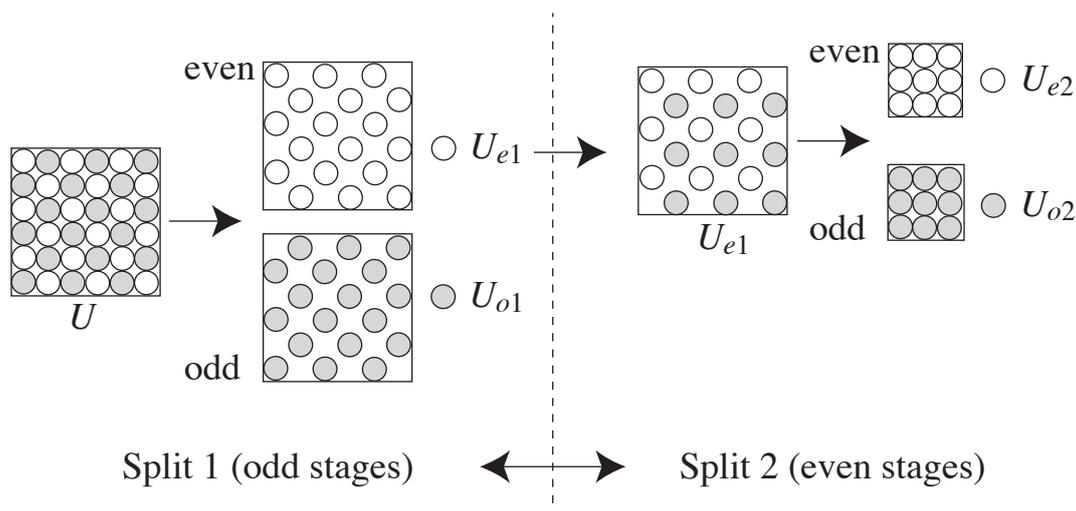


図2 Quincunxサブサンプリング方式

予測処理では、偶数要素 U_e とCNN予測器を用いて奇数要素 U_o を予測する。ここで、一次視覚野においてニューロンが方位選択性を持つことから、CNN予測器は画像の局所的な方向性に応じて6種類の予測器を画素毎に適応的に切り替えている。このようにして得られた奇数要素の予測値と真の奇数要素の画素値の差分をとり、予測誤差を得る。予測誤差の符号量は、CNN予測器の性能(パラメーター)並びに予測器切り替えパターンに依存する。このため、本方式では、最終的に得られた予測誤差を算術符号化した際の見積もり符号量を最小とするよう各層の予測器のパラメーター並びに予測器切り替えパターンを最適化している。

予測により算出された予測誤差 e をコンテキストモデリングに基づき適応多値算術符号化する。また、コンテキスト分類に使用する特徴量は、符号化対象画素を中心とした一定の範囲内の符号化済み予測誤差を用いて算出するが、符号化階層が高階層の予測誤差ほど予測に用いる画素間の距離が遠く高精度な予測が困難であるため、符号化階層数に応じた予測誤差の重み付けにより、コンテキストモデリングの推定精度の向上を図っている。

以上の処理により得られた符号化ビットストリーム及び付加情報(各層の予測器のパラメーター並びに予測器切り替えパターン)を復号側に送信し、復号側では符号側と同様の処理を行うことで、復号の各段階において異なる解像度の画像が再生でき、復号の最後段階において原画像を再生することが可能である。

3. 画素適応CNN予測器のプロトタイプ設計

3.1 出力テンプレートを持つ離散時間型セルラーニューラルネットワーク

図3に出力テンプレートを持つ離散時間型セルラーニューラルネットワークのブロック図を示

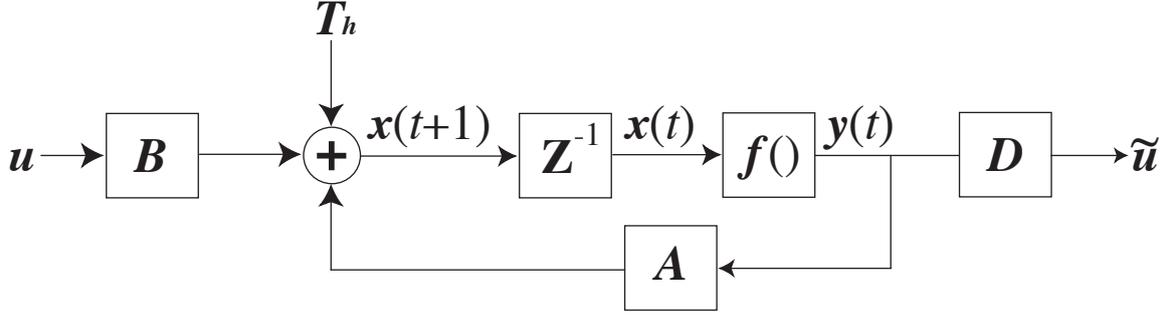


図3 出力テンプレートを持つDT-CNN

す。このニューロンモデルは、人工網膜モデルの一種であるCNNを離散時間化し、ニューロン出力に対して空間フィルタである出力テンプレートを施しているという特徴を持つ。上図の状態方程式は以下の式で与えられる。

$$x_{ij}(t+1) = \sum_{C(k,l) \in N_r(i,j)} B(i,j;k,l)u_{kl} + \sum_{C(k,l) \in N_r(i,j)} A(i,j;k,l)y_{kl} + T_h, \quad (1)$$

$$y_{ij}(t) = f(x_{ij}(t)), \quad (2)$$

$$\tilde{u}_{ij} = \sum_{y(k,l) \in N_r(i,j)} D(i,j;k,l)y_{kl}^e, \quad (3)$$

ここで、 $x_{ij}(t)$ 、 $y_{ij}(t)$ 、 y_{kl}^e 、 u_{ij} 、 \tilde{u}_{ij} はそれぞれニューロンの内部状態、ニューロンの出力、ニューラルネットワークの平衡状態におけるニューロンの出力、CNNへの入力である偶数要素成分、CNNの平衡状態の出力に出力テンプレートを施した結果である最終出力を示す。最終出力は処理対象ステージの分割前の入力画像の予測値となっており、これより奇数要素成分の予測値を算出可能である。また、 T_h 、 $A(i,j;k,l)$ 、 $B(i,j;k,l)$ 、 $D(i,j;k,l)$ 、 $f()$ はそれぞれ、閾値、フィードバックテンプレート(Aテンプレート)、フィードフォワードテンプレート(Bテンプレート)、出力テンプレート(Dテンプレート)、活性化関数を示す。活性化関数は区分線形関数であり、その傾きはテンプレートの種類に関わらず出力値が同一となるようAテンプレートの正規化パラメーターに基づき決定される。ニューロン $C(i,j)$ の r 近傍 $N_r(i,j)$ は $N_r(i,j) = \{C(k,l) \mid \max(|k-i|, |l-j|) \leq r\}$ として定義される。

3.2 画素適応CNN予測器のプロトタイプ設計

一次視覚野においてニューロンが方位選択性を持つことから、CNN予測器は画像の局所的な方向性に応じて6種類の予測器を画素毎に適応的に切り替えている。また、6種類の予測器は、一次視覚野での単純細胞のモデルである2次元ガボールフィルタを単純化した2次元ガウス関数に対し、回転や膨張を加えることにより以下のように設計される。

$$A_{odd} = A_{odd}(i, j; k, l), \quad C(k, l) \in N_r(i, j) \quad (4)$$

$$= \begin{cases} -\alpha_{odd} \exp\left(-\frac{k'^2}{2\sigma_x^2} - \frac{\gamma^2 l'^2}{2\sigma_y^2}\right) & \text{if } (k+l) \bmod 2 = 0 \text{ and } (k, l) \neq (i, j), \\ 0 & \text{otherwise,} \end{cases}$$

$$A_{even} = A_{even}(i, j; k, l), \quad C(k, l) \in N_r(i, j) \quad (5)$$

$$= \begin{cases} 0 & \text{if } k = i \text{ and } l = j, \\ -\alpha_{even} \exp\left(-\frac{k'^2}{2\sigma_x^2} - \frac{\gamma^2 l'^2}{2\sigma_y^2}\right) & \text{otherwise,} \end{cases}$$

$$k' = (k-l) \cos \theta + (l-j) \sin \theta, \quad (6)$$

$$l' = -(k-l) \sin \theta + (l-j) \cos \theta, \quad (7)$$

$$B = B(i, j; k, l), \quad C(k, l) \in N_r(i, j) \quad (8)$$

$$= \begin{cases} 1 & \text{if } k = i \text{ and } l = j, \\ 0 & \text{otherwise,} \end{cases}$$

$$T_h = O, \quad (9)$$

ここで A_{odd} , A_{even} , σ_x , σ_y , γ , θ , O はそれぞれ、奇数ステージのAテンプレートのプロトタイプ、偶数ステージのAテンプレートのプロトタイプ、ガウス関数の水平方向の分散、ガウス関数の垂直方向の分散、ガウス関数の縦横比、ガウス関数の回転角、および零行列を示す。なお、 α_{odd} , α_{even} は各Aテンプレートの係数和を1とするための正規化パラメーターであり、次式で定義される。

$$\alpha_{odd} = \frac{1}{\sum_{(k,l)} A_{odd}(i, j; k, l)}, \quad (10)$$

$$\alpha_{even} = \frac{1}{\sum_{(k,l)} A_{even}(i, j; k, l)}. \quad (11)$$

画素適応予測は、上述の6種のテンプレートを画像の局所的な方向性に応じて切り替えることにより実現されている。本方式では、偶数要素画像から、符号化対象ステージの入力画像を線形予測により簡易的に補間した画像に対し、ソーベルフィルタを適用し、方向性を解析することにより図4のようなテンプレート選択マップを作成している。6種のテンプレートは図4の各領域に応じて画素毎に切り替えられる。各領域を定義する6種の閾値 Th_i については、予測器と同様に、実際に算術符号化した際の符号量が最小となるように学習される。なお、復号側でも符号側の偶数要素画像を持つため、付加情報無しにテンプレートマップを利用することができる。

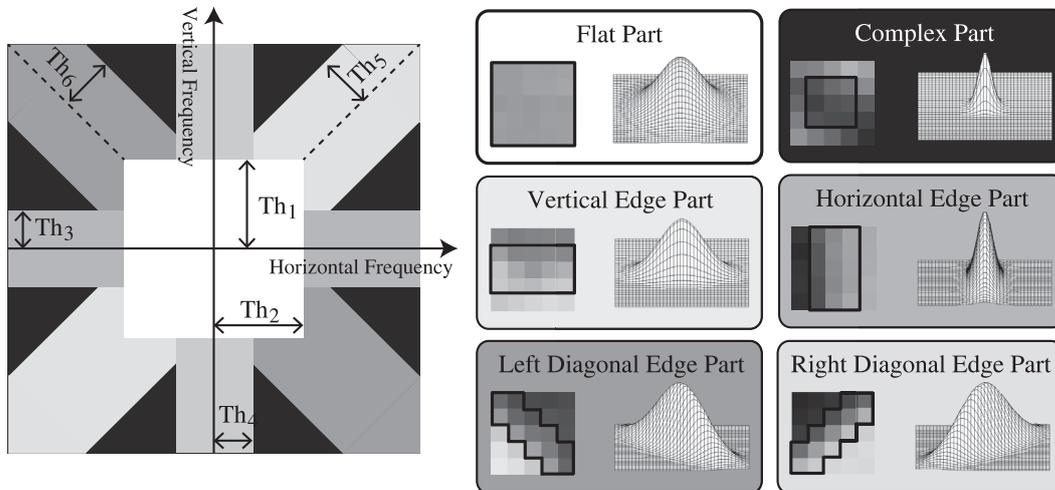


図4 6種類のテンプレートの適用範囲を決定するテンプレートマップ

奇数要素画像はDT-CNNのダイナミクスの平衡状態の出力に対し、ステージ毎に最適化される出力テンプレートを施すことにより得られる。奇数ステージでの出力テンプレートは次式で表現される。

$$D_{odd} = D_{odd}(i, j; k, l), \quad C(k, l) \in N_r(i, j) \quad (12)$$

$$= \begin{cases} -\beta_{odd} \exp\left(-\frac{k'^2}{2\sigma_x^2} - \frac{\gamma^2 l'^2}{2\sigma_y^2}\right) & \text{if } (k+l) \bmod 2 = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

偶数ステージでも同様に、奇数ステージのテンプレートを空間的に拡張した出力テンプレートが用いられ、次式で定義される。

$$D_{even} = D_{even}(i, j; k, l), \quad C(k, l) \in N'_r(i, j) \quad (14)$$

$$= \begin{cases} -\beta_{even} \exp\left(-\frac{k'^2}{2\sigma_x^2} - \frac{\gamma^2 l'^2}{2\sigma_y^2}\right) & \text{if } (i+k) \bmod 2 = 1 \text{ and } (j+l) \bmod 2 = 1, \\ 0 & \text{otherwise,} \end{cases} \quad (15)$$

ここで、 $N'_r(i, j) = \{C(k, l) \mid \max(|k-i|, |l-j|) \leq rd_m\}$ である。出力テンプレートも各ステージにより係数和が異なるため、パラメータ β_{odd} , β_{even} により正規化を行う。

$$\beta_{odd} = \frac{1}{\sum_{(k,l)} D_{odd}(i, j; k, l)}, \quad (16)$$

$$\beta_{even} = \frac{1}{\sum_{(k,l)} D_{even}(i, j; k, l)}. \quad (17)$$

3.3 コンテキストモデリングを用いた予測誤差のエントロピー符号化

エントロピー符号化部では、各ステージの予測誤差 e をコンテキストモデリングに基づいた適応算術符号により符号化する。コンテキストモデリングは、符号化済みの周辺画素群の状態から注目画素における予測誤差の確率密度関数 $P(e)$ を推定する手法であり、本方式では下式で表される特徴量 $U(e)$ を用いて推定している。

$$U(e) = \sum_{k=1}^N \frac{1}{\delta_k} \frac{1}{\delta_s + 1} |e_{d,k}|, \quad (18)$$

ここで、 δ_k は注目画素 P_0 と参照画素 P_k 間のマンハッタン距離である。従って、特徴量 U は周辺画素の予測誤差の絶対値和を注目画素と参照画素の距離の逆数で重み付けし、さらに注目画素と参照画素の間の符号化階層数の差 δ_s に応じた重み付けをしている。これより、特徴量 U は周辺画素における予測の正確さを示す指標と見なせ、予測誤差 e の確率密度関数 $P(e)$ の分布形は U の値に応じて変化していると考えられる。そこで、図5に示すように特徴量 U を閾値により32レベルに量子化し、量子化レベルそれぞれに条件付確率密度関数 $P(e|n)$ をモデル化する。各画素のコンテキストは符号化済みの予測誤差と閾値から一意に決定できるため、予測誤差 e の条件付確率モデルをコンテキストに応じて画素単位で切り換える適応エントロピー符号化が可能となる。

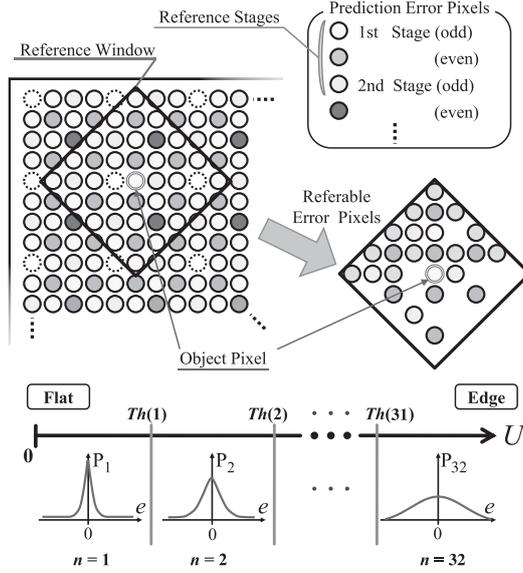


図5 コンテキストモデリングを用いた特徴量の算出

3.3.1 予測誤差の情報量

本方式では、予測誤差の確率密度関数をコンテキスト毎に正確にモデル化するために、次式で示される一般化ガウス関数に基づいた確率モデルを導入する。

$$\tilde{P}(e|n) = \frac{c_n \eta(c_n, \sigma_n)}{2\Gamma(1/c_n)} \exp\{-|\eta(c_n, \sigma_n)e|^{c_n}\}, \quad (19)$$

$$\eta(c_n, \sigma_n) = \frac{1}{\sigma_n} \sqrt{\frac{\Gamma(3/c_n)}{\Gamma(1/c_n)}}, \quad (20)$$

ここで、 $\Gamma(\cdot)$ はガンマ関数、 c_n は確率密度関数の形状を決定するパラメータである。先行研究からの知見より、本パラメータの取る値を0.2~3.2(0.2刻み)の範囲と限定でき、実際の頻度分布に最も合致する値を量子化レベル毎に設定すれば正確なモデリングが可能となるが、実際には符号化効率を考慮する必要があるため、各確率モデル $\tilde{P}(e|n)$ に基づき実際にエントロピー符号化を適用した際に符号化効率が最大となる c_n の値を採用する。また、パラメータ σ_n については先行研究の実験結果より、コンテキスト毎に固有の値を使用するが、特徴量を量子化する閾値の値を可変とすることにより、最適なモデル化を可能にしている。また、計算機での演算における丸め誤差に起因する影響を防ぐため、予測値 p をステップ幅 h_s で線形量子化している。例えば、8ビット精度画像では、輝度値は0~255の整数値で表現されるため、 p が既知であれば予測誤差 e の取り得る範囲は

$$e \in \{s - p \mid s = 0, 1, \dots, 255\} \quad (21)$$

と制限できる。以上より、予測値 p とコンテキスト n が与えられたときの予測誤差 e の条件付生起確率を確率モデル $\tilde{P}(e|n)$ を用いて次式のように定義することができる。

$$P_r(e|p, n) = \frac{P_r(e|n)}{\sum_{s=0}^{255} P_r(s - p|n)}, \quad (22)$$

$$P_r(e|n) = \int_{-h_s/2}^{h_s/2} \tilde{P}(e + \varepsilon|n) d\varepsilon, \quad (23)$$

これより、適応算術符号化を適用した際の各画素における予測誤差の符号化量を次式によって見積ることができる。

$$L(e|p, n) = -\log_2 P_r(e|p, n) = \log_2 \left\{ \sum_{s=0}^{255} P_r(s - p|n) \right\} - \log_2 P_r(e|n) \quad (24)$$

3.4 PSOを用いた見積符号量最小化学習

本方式の符号化性能は各ステージのCNN予測器の予測性能に大きく依存しているため、CNN予測器の予測性能を決定づけるパラメーター並びにテンプレート選択マップの閾値を適切に決定することが肝要である。本研究では、最終階層から再生順に予測器を反復的に最適化することで、予測誤差を算術符号化時の見積もりの符号量を評価することを可能としている。さらに、最適化の手法として、目的関数の微分係数を必要としないメタヒューリスティックな最適化手法であるPSOを学習に用いることにより、学習の困難さに対処している。ここで、算術符号化における予測誤差の見積符号量を次式で定義し、これをPSOで評価する関数として予測器設計を行うことで実際に符号化した際の符号量を削減することが可能となる。

$$\text{cost}(\sigma, \mathbf{Th}) \equiv -\log_2 \Pr(e | \hat{s}(p), n), \quad (25)$$

ここで、 $\mathbf{Th} = \{Th_1, Th_2, \dots, Th_6\}$, $\sigma = \{\sigma_f, \sigma_c, \sigma_{lv}, \sigma_{sv}, \sigma_{sd}, \sigma_{ld}\}$ はそれぞれ、テンプレート選択マップの各閾値、方向性を考慮したテンプレートの形状を決定するガウス関数の標準偏差を示す。

4. 実験結果

提案方式の符号化性能を客観的に評価するため、図6に示す10種類の8ビット精度標準画像に対し、代表的な可逆符号化方式との比較を行った。符号化に用いるパラメーターはPSOを用いた学習により画像毎に最適な値を学習により求め、それを付加情報とするが、PSOにより求めることが困難なパラメーターは、事前実験により求めた最適な値を符号化側と復号側で共通の値を用いている(符号化のレベル数 $L=6$ (12ステージ), 平坦領域用テンプレートサイズ 5×5 , 複雑な領域用テンプレートサイズ 3×3 , その他の領域用テンプレートサイズ 3×5)。

提案方式と国際標準方式の符号化性能を比較した結果を表1に示す。ここで、JPEG-LS[4]およびJPEG2000[5]は静止画像符号化の国際標準方式である。

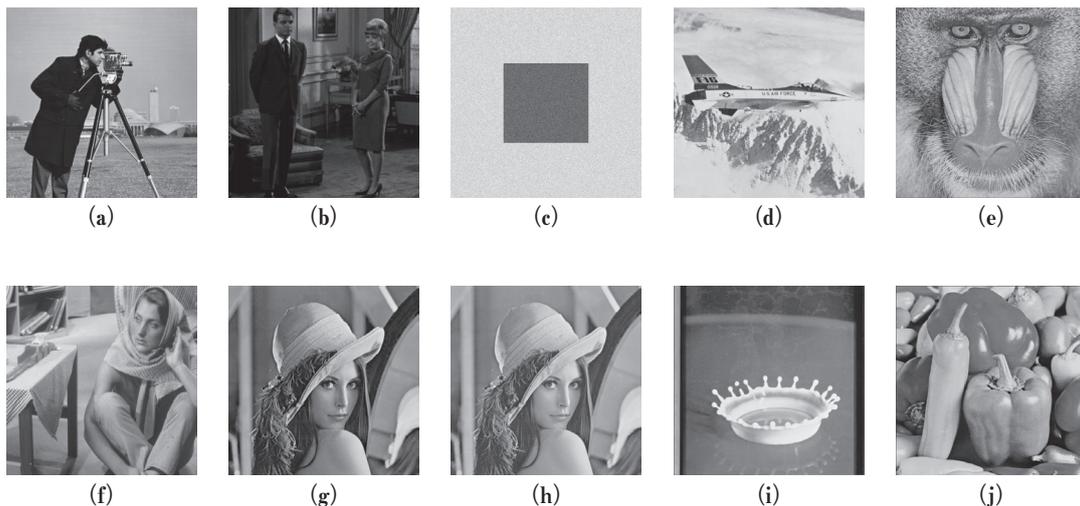


図6 符号化対象画像 : (a) Camera, (b) Couple, (c) Noisesquare, (d) Airplane, (e) Baboon, (f) Barbara, (g) Lena, (h) Lenagray, (i) Milkdrop, (j) Peppers

表1 符号化レートの比較[bits/pel]

Image	提案方式	JPEG2000	JPEG-LS
camera	4.325	4.540	4.314
couple	3.735	3.919	3.699
noisesquare	5.413	5.639	5.683
airplane	3.746	4.015	3.817
baboon	5.870	6.109	6.036
barbara	4.688	4.845	4.906
lena	4.449	4.685	4.607
lennagray	4.068	4.306	4.238
milkdrop	3.564	3.768	3.630
peppers	4.460	4.631	4.513
<i>Average</i>	4.432	4.646	4.544

本方式の符号化レートはJPEG2000およびJPEG-LSの符号化レートに比べて平均でそれぞれ0.234 [bits/pel]および0.112[bits/pel]低い値を達成しており、本方式の有効性を確認した。

5. 結論

本論文では、様々な再生機器解像度に対応するデジタルアーカイブ構築のための高能率画像符号化手法として、画像のエッジ方向性に応じた6種類のテンプレート画素毎に最適選択するCNNを用いた階層的な可逆符号化を提案した。提案方式は、各階層において予測に用いられる奇数要素画像から、エッジの方向性を解析しテンプレートマップを求め、このマップ情報に基づき6種類のテンプレート画素毎に最適選択している。各階層において、CNN予測器のパラメータ及びテンプレート選択マップの閾値は、画像毎に実際の符号化した際の見積もり符号量を最小化するようにPSOを用いて再生順に学習される。

提案方式の有効性は、10種類の標準画像に対するコンピュータシミュレーションを行い確認し、国際標準であるJPEG-LS, JPEG2000と比較し、高い符号化効率を持つことを明らかにした。

参考文献

- [1] L. O. Chua and L. Yang, “Cellular neural networks: Theory,” IEEE Trans. Circuits and Syst., vol. 35, no. 10, pp. 1257–1272, oct. 1988.
- [2] J. Kennedy and R. Eberhart, “Particle Swarm Optimization,” Proc. The 1995 IEEE International Conference on Neural Networks, vol. IV, pp.1942-1948, 1995.
- [3] S. Takenouchi, H. Aomori, T. Otake, M. Tanaka, I. Matsuda, and S. Itoh, “Hierarchical lossless image coding using cellular neural network,” International Conference on Neural Information Processing (ICONIP2010), pp. 679–686, nov. 2010.
- [4] ISO/IEC 14495-1:1999, “Information technology — lossless and near lossless compression of continuous-tone still images: Baseline,” dec. 1999.
- [5] ISO/IEC 15444-1:2004, “Information technology — jpeg 2000 image coding system part 1: Core coding system,” dec. 2000.