Scalable Lossless Image Compression Method Using CNN Predictors with Estimated Coding Bits Minimization Learning

Hisashi Aomori*, Keisuke Takizawa†, Tsuyoshi Otake‡, Masatoshi Sato§,
Ichiro Matsuda†, Susumu Itoh† and Mamoru Tanaka¶

* Department of Electrical and Electronic Engineering, Chukyo University
101–2 Yagoto-honmachi, Showa-ku, Nagoya, Aichi, 466–8666 Japan
†Department of Electrical Engineering, Tokyo University of Science
2641 Yamazaki, Noda-shi, Chiba, 278–8510 Japan
‡Department of Software Science, Tamagawa University
6–1–1, Tamagawa-Gakuen, Machida, Tokyo, 194–8610 Japan
§Division of Management Systems Engineering, Tokyo Metropolitan University
6–6 Asahigaoka, Hino, Tokyo 191-0065, Japan
¶Faculty of Science and Technology, Sophia University
7–1, Kioi-cho, Chiyoda-ku, Tokyo 102–8554 Japan
Email: aomori@ieee.org

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. Introduction

Digital archiving applications are receiving a lot of attention now. They were mainly used for archiving artistical legacies, however, the utilization of them is drastically changed because of the rapid development of electronics devices and demand of counter plan against natural disaster. Digital archiving applications such as legacy-archive, artwork-archive, disaster-archive and so on, are demanded to be used by various devices such as PC, smartphone, tablet PC and etc. Therefore, to realize various reproduction resolution, the scalable image coding scheme is indispensable for modern digital archiving applications. Also, lossless image coding schemes are required for digital archiving.

In this paper, we propose a novel scalable lossless image coding scheme with pel-adaptive prediction using cellular neural network (CNN). In the field of the high efficiency lossless image coding, a pel-adaptive predictor enables high prediction performance. However, in general, huge additional information for predictor selection spoils this advantage. To deal with this difficulty, we use edge-orientation predictors consist of space-variant CNN for pel-adaptive prediction. The advantage of this method is that since predictors are selected based on the edge information, pel-adaptive prediction can be realized without selection information. Moreover, every parameters of CNN predictors are optimized by the estimated coding rate minimization learning. Since the cost function has very complex shapes and many parameters must be optimized, this learning processes are very hard problems. To overcome this problem, a greedy algorithm based learning is used for obtaining initial parameters of particle swarm optimization (PSO) based learning.

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.

2. Scalable lossless image coding schme using CNN

The proposed scalable lossless image coding scheme using pel-adaptive CNN predictors is shown in Fig. 1. In this system, an input original image U is divided into even polyphase components U_e and odd polyphase components U_o like Fig. 2 left side. Furthermore, divided even polyphase components U_e are divided again like Fig. 2 right side to realize a hierarchical coding scheme. As shown in Fig. 2, there are two image dividing patterns, and they are called odd stage and even stage. Moreover, a set of odd and even stages is called one level. Therefore, this hierarchical coding scheme enables multi-resolution representation of the input image, and scalable coding bitstream can be obtained.

The odd components U_o are predicted from the even components U_e via the pel-adaptive CNN predictors that are selected by edge-orientation of the even components. In this work, the six types of CNN prototype predictors are prepared and total 12 parameters of CNN predictor are decided in order to minimize the actual coding rate. The CNN parameters, which determine the distribution characteristic of the prediction error, are learned to minimize an objective function that evaluates the actual coding rate.

The prediction errors e are encoded by a multi-level arithmetic coding with a context modeling where distributions of prediction errors are modeled by the sixteen types of PDFs given by generalized Gaussian functions. Finally, encoded bit stream and side information are transmitted to the decoder. In the decoder, reversible processes of the encoder are applied, and the input image can be reconstructed without any loss.

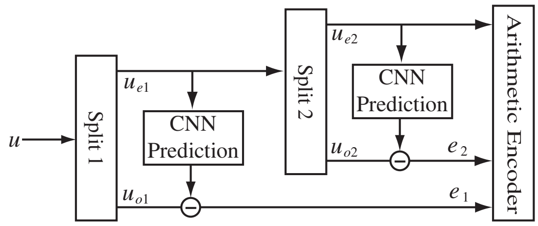

Figure 1: The hierarchical encoder using CNN predictors

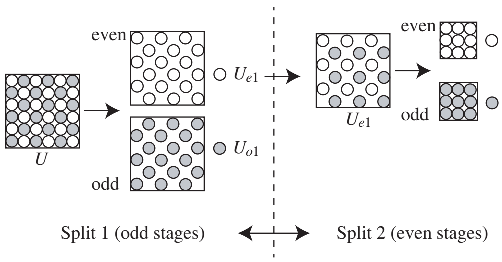

Figure 2: Quincunx sampling

2.1. The prototype design of pel-adaptive CNN predictors

2.1.1. Discrete time CNN with output template

The block diagram of the discrete time CNN (DT-CNN) with an output filter is illustrated in Fig. 3, and its state equation is described as

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

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

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

where $x_{ij}(t)$, $y_{ij}(t)$, y^e_{ij}, u_{ij}, and $\tilde{u}_{ij}$ are the internal state, the output, the equilibrium output, equilibrium solution for even components, the input of the cell (basic processing unit of CNN), the final output, respectively. Also, T_h, $A(i,j;k,l)$, $B(i,j;k,l)$, and $D(i,j;k,l)$ are the threshold, the feed-back template, the feed-forward template, and the output template, respectively. The function $f()$ is a piece wise linear (PWL) output function and its slope is determined by the sum of A-template coefficients. The r-neighborhood $N_r(i,j)$ is defined by $N_r(i,j) = \{C(k,l) | \max\{|k-i|, |l-j|\} \le r\}$.

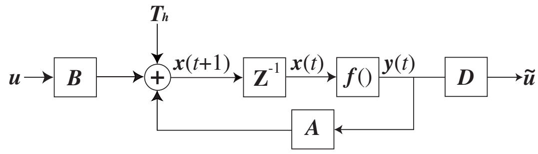

Figure 3: DT-CNN with output template

2.1.2. The prototype design of pel-adaptive CNN predictors

From the findings of our previous work [1], the prototype of edge-orientation templates are given by

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

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

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

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

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

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

$$\boldsymbol{B} = B(i,j;k,l), \qquad C(k,l) \in N_r(i,j) \tag{8}$$

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

$$T_h = O, \tag{9}$$

where $\boldsymbol{A_{odd}}$, $\boldsymbol{A_{even}}$, σ_x, σ_y, γ, θ, O are A template for odd stages, A template for even stages, a standard deviation of horizontal direction, a standard deviation of vertical direction, the spatial aspect ratio, a rotation angle of Gaussian function, and the zero matrix, respectively. α_{even} and α_{odd}

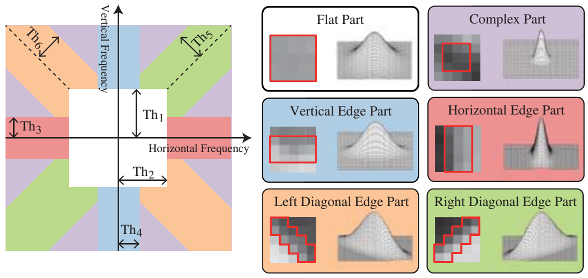

Figure 4: Template switching map in spatial frequency domain: each region range of given image is adaptively decided to minimize the actual coding bits.

are the normalization coefficient given by

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

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

To perform pel-adaptive prediction, configurations of the CNN predictor are adaptively adjusted to the local structure of an input image. In our method, by according to the edge analysis result, an input image is classified into six regions as shown in Fig. 4. Note that the edge analysis image is created from only the even components and a linear filter is used for generating a coarse-interpolated image for this process. Therefore, pel-adaptive prediction with no selection information can be realized.

Then the predicted image of odd components $\tilde{u}$ can be obtained by using the equilibrium solution of the DT-CNNs through output template $\boldsymbol{D}$ that is defined in each stages. In the odd stages, the predicted odd component value in coordinate (i, j) can be given by equilibrium output via the output template $\boldsymbol{D}$. In this case, the coordinate of odd component satisfies $(i + j) \bmod 2 = 1$. The output template for odd stages $\boldsymbol{D}_{odd}$ is defined by

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

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

Also, in the even stages, the predicted odd pixel value in coordinate (i, j) can be given by equilibrium solution via output template $\boldsymbol{D}_{even}$. In this case, the coordinate of odd pixel satisfies $i \bmod 2 = 1$ and $j \bmod 2 = 1$. The $\boldsymbol{D}_{even}$ is spatially extended version of $\boldsymbol{D_{odd}}$ as

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

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

where $N'_r(i,j) = \{\, C(k,l) | max\{|k-i|, |l-j|\,\} \le rd_m\}$. β_{odd} and β_{even} are the normalization coefficient defined by

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

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

2.2. Entropy coding of prediction error

In this layer, prediction error e of each stage is encoded by the adaptive arithmetic coding based on the context modeling. The context of target pixel is given by the context function $U(e)$ degfined by the weighted sum of absolute prediction errors of already encoded pixels.

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

where δ_k is the Manhattan distance between the target pixel and each already encoded pixel in the reference window, δ_s is the hierarchical distance between the target pixel and each already encoded pixel in the reference window, d is the number of reference stages, and N is the number of reference pixels. d and N are experimentally-obtained parameter. In this paper, $d = 3$ (current stage + next one level) and $N = 40$ are used. For efficient encoding, the context should be quantized into sixteen regions by the thresholds $Th(n)$ $(n = 1, 2, \ldots, 15)$ which are optimized for minimizing the actual coding rate (see Fig. 5).

2.3. Coding rate minimization learning using PSO

To achieve pel-adaptive prediction and high efficient coding performance, the optimal CNN parameters: a standard deviation of Gaussian function in each template, and thresholds in template selection map, should be decided. In encoder, since the odd polyphase components which are the desire output of predictor can be used for parameter learning, by using the odd polyphase components as a supervised data, all parameters in the encoder can be learned. Then we introduce the cost function for supervised learning that evaluates the actual coding bits of prediction error as

$$\mathrm{cost}(\boldsymbol{\sigma}, \boldsymbol{Th}) \equiv -\log_2 \Pr(e \,|\, \hat{s}(p), n), \tag{19}$$

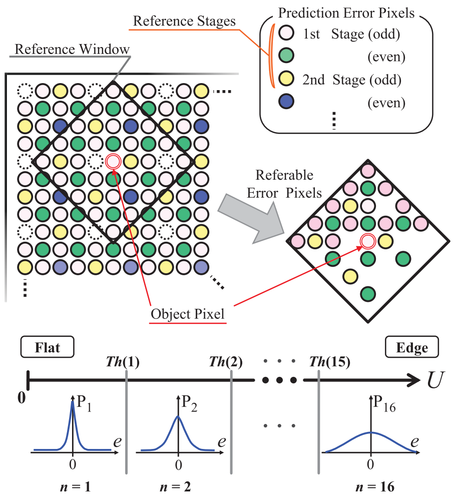

Figure 5: Calculating the feature quantity U and the context modeling.

where $\boldsymbol{Th} = (Th_1, Th_2, Th_3, Th_4, Th_5, Th_6)$ is the threshold vector, and $\boldsymbol{\sigma} = (\sigma_f, \sigma_c, \sigma_{lv}, \sigma_{sv}, \sigma_{ld}, \sigma_{sd})$ is the standard deviation vector: σ_f is the standard deviation of flat part, σ_c is the standard deviation of complex part, σ_{lv} is the long-side-direction standard deviation of vertical/horizontal edge parts, σ_{sv} is the short-side-direction standard deviation of vertical/horizontal edge parts, σ_{ld} is the long-side-direction standard deviation of diagonal edge parts, and σ_{sd} is the short-side-direction standard deviation of diagonal edge parts, respectively. Since to determine a context of prediction error that gives the PDF model, prediction errors in higher stages are required, the information flow of learning process is a reverse version of the encoding step. Then we can evaluate a cost of learning process which is defined by (19) using prediction error of higher stages. Since this learning process is very complex and difficult, we first use semi-full search algorithm [2] for obtaining initial parameters of PSO-based learning. Then PSO-based learning [3] is applied to derive optimal parameters which minimize the cost function cost($\boldsymbol{\sigma}$, $\boldsymbol{Th}$).

3. Experimental Results

The coding performance of the proposed method was compared with that of the JPEG 2000 and the JPEG-LS. The coding parameters of proposed system are decided by pre-experiments: the number of level $L = 6$ (12 stages), template size is 5×5 for flat and edge part, and 3×5 is for other regions.

Table 1 lists the coding performance of each method. The average coding rate of the proposed method achieved 0.234, and 0.112 bits/pel lower than those of the JPEG2000 standard, and JPEG-LS standard, respectively. This results suggest that our proposed image coding framework has excellent coding performance.

Table 1: Lossless coding rate [bits/pel]

Image	Proposed	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
Average	**4.432**	4.646	4.544

4. Conclusion

In this paper, we have propose a novel scalable lossless image coding scheme with pel-adaptive prediction using CNN. The edge-orientation map obtained by only even polyphase components enables a pel-adaptive prediction with no selection information. Moreover, the PSO-based learning with the initial parameters given by a greedy-algorithm-based learning makes it possible to achieve a high coding efficiency. Experimental results show that our proposed framework is highly efficient for lossless image coding compared with conventional methods including JPEG standards.

Acknowledgment

This research was partially supported by a grant from the hibi science & technology foundation.

References

[1] K. Takizawa, S. Takenouchi, H. Aomori, T. Otake, M. Tanaka, I. Matsuda, and S. Itoh, "Lossless image coding by cellulear neural networks with minimum coding rate learning," *Proceedings of the 2011 European Conference on Circuit Theory and Design (ECCTD2011)*, pp. 13 – 16, aug 2011.

[2] 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.

[3] K. Takizawa, S. Takenouchi, H. Aomori, T. Otake, M. Tanaka, I. Matsuda, and S. Itoh, "Coding rate minimization learning method for cnn image predictors in lossless image coding scheme by particle swarm optimization," *IEICE Technical Report*, vol. 111, no. 106, pp. 119 – 124, jul 2011.