NOLTA, IEICE

Paper

Progressive image transmission based on image spatio-temporal decomposition by sigma-delta cellular neural network

Hisashi Aomori [1 a)], *Ryohei Mizutani* [2], *Hideharu Toda* [1], and *Tsuyoshi Otake* [3]

[1] *Faculty of Engineering, Chukyo University*
101–2 Yagoto-honmachi, Showa-ku, Nagoya, Aichi, 466–8666, Japan

[2] *Yamazaki Mazak Corporation*
1–131 Takeda, Oguchi-cho, Niwa-gun, Aichi, 480–0197, Japan

[3] *College of Engineering, Tamagawa University*
6–1 Tamagawagakuen, Machida-shi, Tokyo, 194–0041, Japan

[a)] *aomori@ieee.org*

Received October 17, 2021; Revised December 16, 2021; Published April 1, 2022

Abstract: In this paper, we propose a novel progressive image transmission framework based on spatio-temporal image decomposition and synthesis by the SD-CNN. In our method, we redesign the baseline SD-CNN and the weighted sum is introduced in the accumulator. This innovation enables lossless or near-lossless progressive image transmission. Experimental results in various test images support that the image reconstruction performance of the SD-CNN has dramatically improved by our method.

Key Words: spatio-temporal image decomposition, cellular neural network, progressive image transmission

1. Introduction

The progressive image transmission technique is widely used in remote sensing, medical images, and telemedicine applications. In the progressive image transmission, an input image is transformed into an image sequence. During reconstruction, a coarse approximation is firstly reconstructed by receiving a bit stream of early stages, then the quality of the reconstructed image is gradually enhanced by adding subsequent sequence. Also, progressive image transmission is often desired for band-limited communication channels since it can reduce the number of bits per transmission. Traditional progressive image transmission can be categorized into two approaches [1–4]. One is a multi-resolution or hierarchical coding method with the discrete wavelet transform (DWT). Most of this framework incorporates image compression methods. Especially, the JPEG 2000 and the SPIHT [5] are well-known implementations. The other is the bit-plane method (BPM) that is the simplest implementation of a progressive image transmission. This method utilizes the bit-plane decomposition. A n-bit im-

  DOI: 10.1587/nolta.13.264

age is decomposed into n bit-planes, where i-th bit-plane is given by the i-th bit of binary numbers representing a pixel value. Meanwhile, a few bio-inspired approaches have been studied. [6] is a well-known implementation, where radial basis function (RBF) networks are used as an interpolator for image coding scheme. A majority of bio-inspired approaches are used in image coding frameworks, while only a few methods achieve progressive image transmission framework utilizing a nature of the biological system.

In the visual system, light is captured by the eyes, and its luminance value is converted to spatio-temporal spikes by the retina. This functionality suggests that the visual system has the potential of spatio-temporal image decomposition which is applicable to progressive image transmission. In [7], we proposed the sigma-delta cellular neural network (SD-CNN) which enables a spatial domain sigma-delta modulation. Like a sigma-delta modulation (SDM), a luminance value of each pixel is represented by pulse density modulated bitstream. Since the CNN is also a mimic of the retina, the essential functionality of the SD-CNN can be thought of a spatio-temporal decomposition of an image. We experimentally showed that the SD-CNN can decompose a still image into a binary image sequence and the quality of the reconstructed image is gradually enhanced by adding binary image sequence. However, the development of the SD-CNN model that achieves high fidelity progressive image transmission still remains an open research issue.

In this paper, we propose a spatio-temporal image decomposition for progressive image transmission by the SD-CNN. In our method, from the findings of the image compression framework using CNNs [8], we redesign the SD-CNN to improve image reconstruction performance. This innovation has dramatically improved the image reconstruction performance and enabled lossless or near-lossless progressive image transmission. Furthermore, by taking the weighted sum of an output binary sequence, we improve the efficiency of the accumulator which recovers the input image from a binary image sequence output by the SD-CNN. Since progressive image transmission by our method consists of pulse density modulation of an image and its demodulation, the mechanism of our approach is completely different from conventional methods which utilize the DWT and the BPM. The encoding and decoding processes for progressive image transmission in the proposed method are as follows. In the encoder, an input image is pulse density modulated by the SD-CNN, and we can obtain one binary image at each iteration of the SD-CNN dynamics. Then the output binary image sequence is transmitted to the decoder. In the decoder, the input image is restored via the accumulator, where the weighted sum of each image in the binary image sequence is computed. This image recovery process is inspired by the composition process of the BPM. Experiments on various grayscale test images confirm that the proposed method has lossless or near-lossless progressive image transmission performance.

The rest of this paper is organized as follows. In section 2, a novel high fidelity progressive image transmission framework using the SD-CNN is proposed. Details of redesigning the SD-CNN is given, and the definition of the new accumulator used in our method is described. Experiments on progressive image transmission performance and efficiency are presented in section 3. Finally, some concluding remarks are made in section 4.

2. Spatio-temporal image decomposition and progressive recovery by sigma-delta cellular neural network

Figure 1 shows the encoder and the decoder of our proposed progressive image transmission framework. A cell which is a processing unit of CNN and a pixel of the input image have a one-to-one correspondence. Each cell is locally connected to its neighbors called r-neighborhood indicated by the red rectangle in Fig. 1. Also, pixel values are converted to real numbers between -1 and 1 that are corresponding to the dynamic range of CNN state variables.

In the encoder, an input image is decomposed into a spatio-temporal binary image sequence by the dynamics of the SD-CNN. As shown in Fig. 2, we can obtain one binary image at each iteration of the SD-CNN dynamics. In our method, we utilize the findings of CNN-based image compression framework [8] and the SD-CNN is redesigned for high fidelity image reconstruction. The first few output images of the proposed method are very similar to that of the BPM indicated in Fig. 3. The output sequence is transmitted to the decoder.

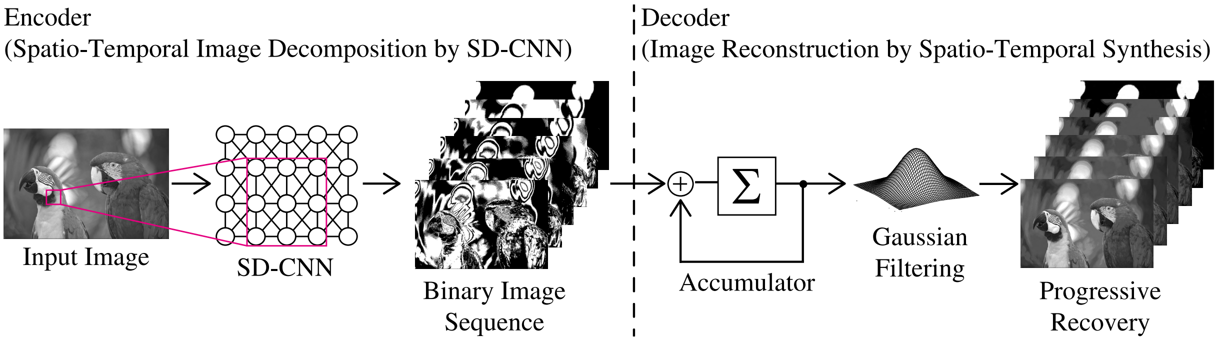


Fig. 1: Proposed system

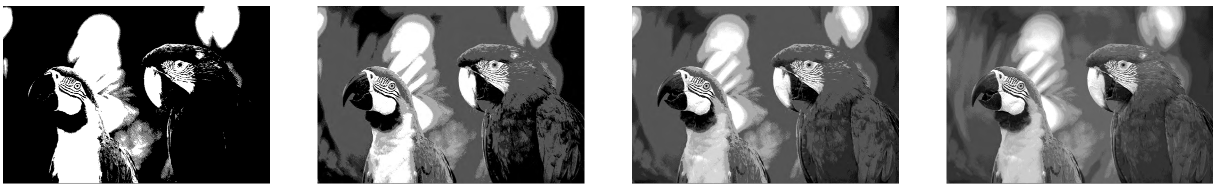
(a) 1st output (b) 2nd output (c) 3rd output (d) 4th output

Fig. 2: Example of output sequence of our method

(a) 8th plane (b) 7th plane (c) 6th plane (d) 5th plane

Fig. 3: Example of output sequence of the BPM

In the decoder, the original image is recovered progressively shown in Fig. 4 from a binary image sequence by spatio-temporal synthesis. The processes of spatio-temporal synthesis is very simple as described below. Each binary image becomes an input to the accumulator and its weighted sum is calculated. To determine the weight for each binary image, we utilize the weighting of the bit-plane restoration as a reference. That is, they are set so that the first output has the heaviest weight and gradually decreases. Finally, the reconstructed image is obtained by applying the Gaussian filter to the weighted sum of the binary image sequence.



(a) 1st recovered image (b) 2nd recovered image (c) 3rd recovered image (d) 4th recovered image

Fig. 4: Example of progressive image recovery by proposed method

2.1 The SD-CNN for high fidelity image reconstruction

The block diagram of a cell of the SD-CNN is illustrated in Fig. 5. The state equation of a cell at the coordinates (i, j) is given by

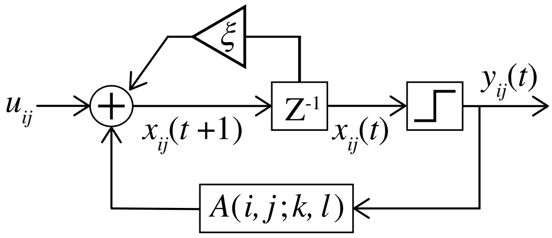


Fig. 5: The block diagram of a cell of the SD-CNN

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

$$y_{ij}(t) = f\left(x_{ij}(t)\right) = \begin{cases} 1 & x_{ij}(t) \geqq 0, \\ -1 & \text{otherwise}, \end{cases} \tag{2}$$

where $x_{ij}(t)$, $y_{ij}(t)$, u_{ij}, $f(\cdot)$, $A(i,j;k,l)$ and ξ are the internal state, the output, the input of a cell, the nonlinear output function, the A-template and the C-template, respectively. The r-neighborhood $N_r(i,j)$ is defined by $N_r(i,j) = \{C(k,l) | \max\{|k-i|, |l-j|\} \leq r\}$. The A-templates is defined by

$$A(i,j;k,l) = -\frac{1}{2\pi\sigma^2} \exp\left(-\frac{(k-i)^2 + (l-j)^2}{2\sigma^2}\right), \tag{3}$$

where σ is a standard deviation of the Gaussian function.

In the baseline SD-CNN [7], the A-template, the C-template, and σ are empirically determined parameters. Especially, an optimal σ depends on an input image. In our method, these parameters are determined by referring to the design policy in [8]:

Design policy (i)

The A-template is designed so that its center $A(i,j;i,j)$ is -1.

Design policy (ii)

The coefficient for $x_{ij}(t)$ is defined to be zero.

Note that the following points regarding the design policy (i) and (ii). Although the convergence condition of the CNN is $A(i,j;i,j) = 0$, in our method, a spatio-temporal image decomposition is achieved by unsatisfying the convergence condition. For this reason, the proposed method adopts design policies. In [8], once the A-template is designed so that its center $A(i,j;i,j)$ is -1, then $A(i,j;i,j)$ is set to 0. The design policy (i) adopts the original design rule. $A(i,j;i,j)$ becomes a coefficient of $x_{ij}(t)$, if $x_{ij}(t) = y_{ij}(t)$. So, the design policy (ii) adopts $x_{ij}(t) = 0$ instead of the convergence condition.

From the design policy (i) and Eq. (3), we get $2\pi\sigma^2 = 1$. This means that σ is independent of an input image. For the design policy (ii), we assume that the output function $f(\cdot)$ is a piecewise linear function defined by $f(x) = \min(1, \max(-1, x))$ and x is in its linear region. Then, the state equation of the SD-CNN is rewritten as

$$x_{ij}(t+1) = \xi x_{ij}(t) + A(i,j;i,j)x_{ij}(t) + \sum_{\substack{C(k,l)\in N_r(i,j) \\ (k,l)\neq(i,j)}} A(i,j;k,l) y_{kl}(t) + u_{ij} \tag{4}$$

$$= \left(\xi + A(i,j;i,j)\right) x_{ij}(t) + \sum_{\substack{C(k,l)\in N_r(i,j) \\ (k,l)\neq(i,j)}} A(i,j;k,l) y_{kl}(t) + u_{ij}. \tag{5}$$

To satisfy the design policy (ii), we have $\xi + A(i,j;i,j) = 0$ from Eq. (5). Therefore, ξ is determined as 1 since $A(i,j;i,j)$ is -1.

2.2 Progressive image reconstruction by spatio-temporal synthesis

The output of the SD-CNN becomes the input of the decoding processes. The spatio-temporal synthesis for image recovery consists of two parts. At the first, the accumulator calculates a weighted sum of the time series of each pixel. To determine the weight for the t-th image, we use the weighting of the bit-plane restoration as a reference. Let B_i be the i-th plane image of the BPM for n bit image, then reconstruction image I is given by $I = \sum_{i=1}^{n} 2^{i-1} B_i$. In our method, we set the weights so that the first output has the heaviest weight and gradually decreases. A weighted sum of the time series of each pixel $s_{ij}(t)$ is given by

$$s_{ij}(t) = \sum_{m=1}^{t} (t + 1 - m) y_{ij}(m) \tag{6}$$

Finally, the reconstructed image at t-th iteration $\tilde{u}_{ij}(t)$ is obtained by applying the D-template.

$$\tilde{u}_{ij}(t) = \sum_{C(k,l)\in N_r(i,j)} D(i,j;k,l) s_{kl}(t), \tag{7}$$

$$D(i,j;k,l) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(k-i)^2 + (l-j)^2}{2\sigma^2}\right). \tag{8}$$

3. Experimental Results

To evaluate the effectiveness of the proposed method, we perform image coding/decoding experiments on standard grayscale images. An example of a coded image sequence and its decoding processes are illustrated in Figs. 2 and 4. In all experiments, the r-neighborhood of the SD-CNN is set to $r = 2$, and the maximum number of iterations is set to 256. The peak signal to noise ratio (PSNR) for an average of the mean square error (MSE) is employed for objective performance criteria. Since some images can be recovered without loss by the proposed method, the average PSNR cannot be calculated (PSNR value for lossless reconstruction is infinity). Hence, PSNR for average MSE is employed for an objective image quality measurement. Generally, if PSNR is higher than 40 dB, two images are perceptually indistinguishable (visually lossless). Moreover, if the PSNR is higher than 50 dB, the distortion between the reconstructed image and the original image reaches a near-lossless quality.

First, the progressive image transmission performance of our method is compared with the baseline SD-CNN [7] and the first-order SDM. A pixel and the SDM have a one-to-one correspondence. Therefore luminance of each pixel becomes an input to the corresponding SDM and an average output of a period implies recovered luminance. In this experiment, we use 9 standard grayscale images indicated in Fig. 6 from the dataset of [7]. Resolution of all images is 512×512.

Fig. 6: Test images from the dataset of [7]

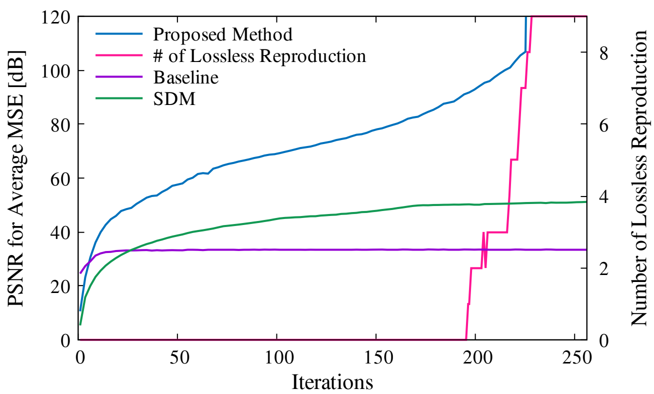


Fig. 7: Progressive image transmission performance of each method

Progressive image transmission performance at each iteration is summarized in Fig. 7. In this figure, the blue curve means the image reconstruction performance of our method, the purple curve is that of the baseline SD-CNN, and the green curve is that of the SDM. The pink curve shows the number of images that are reproduced losslessly by our method. From this figure, we can see that the reconstruction performance and efficiency of our method are much better than that of the baseline SD-CNN and the SDM. In particular, only our method can recover all test images losslessly.

Our method achieves over 40 dB PSNR by only 12 iterations, and the PSNR of the baseline method is always below 40 dB. The proposed method can achieve the same reconstruction performance of the baseline SD-CNN by only 8 iterations. On the other hand, the SDM requires 58 iterations to achieve the same performance as our method.

Next, we evaluate the performance of the proposed method and the SDM on a high-definition image dataset. In this experiment, we use the Kodak photo CD [9] dataset[1] shown in Fig. 8.

Progressive image transmission performance at each iteration is summarized in Fig. 9. In this figure, the blue curve means the image reconstruction performance of our method, and the green curve is that of the SDM. The pink curve shows the number of images that are reproduced losslessly by our method. From this figure, we can see that the progressive image transmission performance improves in proportion to the number of iterations. We also find that not all images can be restored losslessly. The PSNR curve of the SDM towards around 50 dB by 256 iterations, meanwhile, our method can achieve the same performance by only 33 iterations. Therefore, the reconstructed image by the proposed method quickly reaches a near-lossless quality.

Fig. 8: The Kodak image dataset [9]

To evaluate the importance of each image in an image sequence, the starting position of the sequence is varied from 1 to 256 while keeping the sequence length of 256. Experimental results are summarized in Fig. 10 where the blue curve shows image reconstruction performance. We can see that the graph drops rapidly and settles around 50 dB. This indicates that the first output of the SD-CNN has great importance. We believe that this PSNR characteristic is reasonable because the first output of the SD-CNN and the bit plane of the most significant bit in the BPM which plays the most important role for the bit-plane synthesis, are very similar. Also, we can see that if the sequence length is long enough, very high image restoration performance can be obtained.

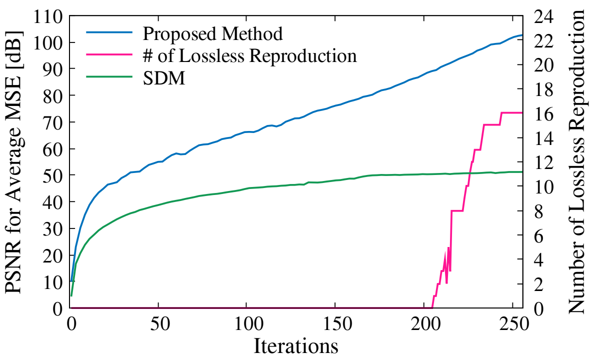


Fig. 9: Progressive image transmission performance for the Kodak dataset

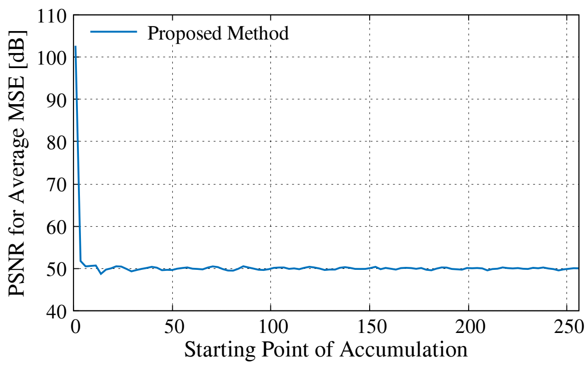


Fig. 10: Evaluation of the importance of each image in an image sequence

[1]This dataset consists of 24 color images and their resolution is 768×512. In this experiment, these images are converted from the RGB color space to the ITU-R BT.601 YCbCr color space, and the Y images are used for the grayscale test images.

4. Conclusion

In this paper, a novel progressive image transmission framework based on spatio-temporal image decomposition and synthesis by the SD-CNN has been proposed. In our method, the baseline SD-CNN is redesigned and the weighted sum is introduced in the accumulator. Experimental results in various test images support that this innovation has dramatically improved the image reconstruction performance and enabled lossless or near-lossless progressive image transmission. It is also found that the first output of the SD-CNN is most significant for image reconstruction.

Acknowledgments

This research was partially supported by a research grant (2020001) of Hagiwara Foundation, and JSPS KAKENHI grant number 21K12052.

References

[1] K. Hung, C. Chang, and I. Lin, "Lossless compression-based progressive image transmission scheme," *The Imaging Science Journal*, vol. 52, no. 4, pp. 212–224, 2004. DOI: 10.1179/136821904225020276

[2] R. Dilmaghani, A. Ahmadian, M. Ghavami, and A. Aghvami, "Progressive medical image transmission and compression," *IEEE Signal Processing Letters*, vol. 11, no. 10, pp. 806–809, 2004. DOI: 10.1109/LSP.2004.835563

[3] J. M. Zheng, D. W. Zhou, and J. L. Geng, "ROI progressive image transmission based on wavelet transform and human visual specialties," *Proc. of ICWAPR'07*, vol. 1, pp. 260–264, 2007. DOI: 10.1109/ICWAPR.2007.4420675

[4] H. Ravikiran and Paramesha, "A hybrid progressive image compression transmission and reconstruction architecture," *Lecture Notes in Electrical Engineering*, vol. 545, pp. 867–875, 2019. DOI: 10.1007/978-981-13-5802-9_75

[5] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996. DOI: 10.1109/76.499834

[6] T. Sigitani, Y. Iiguni, and H. Maeda, "Image interpolation for progressive transmission by using radial basis function networks," *IEEE Trans. Neural Netw.*, vol. 10, no. 2, pp. 381–390, Mar. 1999. DOI: 10.1109/72.750567

[7] H. Aomori, T. Otake, N. Takahashi, and M. Tanaka, "Sigma-delta cellular neural network for 2d modulation," *Neural Networks*, vol. 21, pp. 349–357, Mar. 2008. DOI: 10.1016/j.neunet.2007.12.020

[8] Y. Kawai, H. Toda, H. Aomori, T. Otake, I. Matsuda, and S. Itoh, "Hierarchical lossless image compression using cnn predictors optimized by adaptive differential evolution and context-adaptive coding," *Proc. of NOLTA'20*, pp. 65–67, Nov. 2020.

[9] "Kodak lossless true color image suite," http://r0k.us/graphics/kodak/.